

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 24, 2005, 03:56:10 ; Search time 85 Seconds  
(without alignments)  
2376.054 Million cell updates/sec

Title: US-09-936-759-6

Perfect score: 3001

Sequence: 1 MVRPQRNKKRFFILLNGVWN.....TRDRQPKLVAVHLRLMSEV 563

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2002273 seqs, 358729299 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_23Sep04:\*

- 1: geneseqp1980s:\*
- 2: geneseqp1990s:\*
- 3: geneseqp2000s:\*
- 4: geneseqp2001s:\*
- 5: geneseqp2002s:\*
- 6: geneseqp2003as:\*
- 7: geneseqp2003bs:\*
- 8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2999	99.9	563	3	AB28406	Aab28406 Thermotog
2	932	31.1	602	2	AAW93820	Aaw93820 Bacillus
3	929	31.0	602	2	AAW93825	Aaw93825 Bacillus
4	929	31.0	602	2	AAW93822	Aaw93822 Bacillus
5	929	31.0	602	2	AAW93826	Aaw93826 Bacillus
6	929	31.0	602	3	AB28402	Aab28402 Staphyloc
7	929	31.0	618	2	AAW93821	Aaw93821 Bacillus
8	925	30.8	615	3	AB28408	Aab28408 Codon-opt
9	920	30.7	602	3	AB28409	Aab28409 Salmonell
10	919.5	30.6	598	5	AAW48998	Aam48998 L. gasser
11	902.5	30.1	648	7	AD28924	Adf28924 Murine be
12	899.5	29.9	648	4	AAE02444	Aae02444 Murine be
13	897	29.9	643	4	ABB70164	Abb70164 Drosophil
14	894.5	29.8	648	7	ADD45479	Add45479 Rat Prote
15	894.5	29.8	648	7	AD57446	Ad57446 Rat Prote
16	882	29.4	1010	3	AAW68840	Aaw68840 Fusio
17	879	29.3	603	2	AAW93827	Aaw93827 E. coli G
18	879	29.3	603	2	AAW93824	Aaw93824 Human pr
19	879	29.3	603	3	AB28431	Aab28431 Human bet
20	879	29.3	603	5	AB84108	Abb84108 GUS prote
21	879	29.3	603	6	AB96657	Abp96657 E. coli b
22	879	29.3	603	6	ABJ19649	Abj19649 Artificial
23	877	29.2	618	7	ADJ01666	Adj01666 Modified
24	875.5	29.2	602	2	AAW43387	Aaw43387 Beta-gluc
25	875	29.2	603	5	AB84107	Abb84107 GUS prote

26	875	29.2	604	7	ADD27986	Add27986 Beta-gluc
27	875	29.2	659	7	ADD27989	Add27989 Oleosin/b
28	875	29.2	850	7	ADD27991	Add27991 Caleosin/
29	872.5	29.1	602	2	AAW42429	Aaw42429 Escherich
30	872.5	29.1	711	6	ABR83626	AbR83626 SUMO-beta
31	872.5	29.1	1242	5	ABB81108	Abb81108 LUC-U3'-U
32	872.5	29.1	1242	6	ABB84637	Abb84637 LUC-U3'-U
33	870	29.0	832	2	AAW04302	Aaw04302 Antibody/
34	869.5	29.0	602	1	AAW82948	Aaw82948 Beta-gluc
35	865.5	28.8	613	2	AAW93823	Aaw93823 E. coli G
36	865.5	28.8	613	2	AAW93828	Aaw93828 Human GUS
37	865.5	28.8	613	3	AB28407	Aab28407 Escherich
38	865.5	28.8	633	4	AB62276	Aab62276 Mutant he
39	865.5	28.8	633	4	AB62271	Aab62271 Heavy cha
40	865.5	28.8	651	4	AAE02443	Aae02443 Human bet
41	865.5	28.8	651	7	ADD45481	Add45481 Human pro
42	865.5	28.8	651	7	AD57448	Ad57448 Human pro
43	865.5	28.8	651	8	ADP12392	Adp12392 Protein e
44	865.5	28.8	722	6	AAW33322	Aaw33322 L. mexica
45	865.5	28.8	722	8	ADF47503	Adf47503 Human bet

ALIGNMENTS

RESULT 1	
AB28406	
ID	AB28406 standard; protein; 563 AA.
XX	
AC	AB28406;
XX	
DT	26-JAN-2001 (first entry)
XX	
DE	Thermotoga maritima beta-glucuronidase.
XX	
KW	Microbial; beta-glucuronidase; GUS; Enterobacter; Salmonella;
KW	Pseudomonas; Staphylococcus; Thermotoga; transgenic plant; bioindicator;
KW	transgenic insect; marker; glucuronide detoxification.
XX	
OS	Thermotoga maritima.
XX	
PN	WO200055333-A1.
XX	
PD	21-SEP-2000.
XX	
PF	16-MAR-2000; 2000WO-US007107.
XX	
PR	17-MAR-1999; 99US-00270957.
XX	
PA	(CAMP-) CAMBIA BIOSYSTEMS LLC.
XX	
PI	Jefferson RA, Mayer JE;
XX	
DR	WPI; 2000-647075/62.
XX	
DR	N-PSDB; AAA07937.
XX	
PT	Novel microbial beta-glucuronidase genes and gene products used as
PT	reporter/effector molecule, as diagnostic tool, in positive selection, to
PT	target molecules to specific cells and to detect and track linked genes.
XX	
PS	Claim 3; Fig 5B; 116bp; English.
XX	
CC	The present sequence is a microbial beta-glucuronidase (GUS) protein. GUS
CC	genes were obtained from six different genera: Enterobacter/Salmonella,
CC	Pseudomonas, Salmonella, Staphylococcus and Thermotoga. Microbial GUS can
CC	be used as a reporter/effector molecule for transgenic constructions and
CC	in vitro diagnostic applications. It may also be used to generate
CC	sentinel plants that serve as bioindicators of environmental status. It
CC	may be used to generate transgenic insects for tracking insect
CC	populations or to facilitate the development of a bioassay for compounds
CC	that affect molecules critical for insect development (e.g. juvenile
CC	hormone). Secreted GUS may also serve as a marker for beneficial fungi
CC	destined for release into the environment. In animal systems, secreted

CC GUS may be used to achieve extracellular detoxification of glucuronides  
(e.g. toxin glucuronide) and to examine conjugation patterns of  
CC glucuronides. Microbial GUS may also be used in traditional medical  
CC diagnostic assays, for drug testing, pharmacokinetic studies,  
CC bioavailability studies, diagnosis of diseases and syndromes, following  
CC progression of disease or its response to therapy. Microbial GUS has  
CC increased thermal stability, high turnover number and enzymatic activity.  
CC It is highly specific for the substrate and water soluble, and the  
CC substrates are stable

XX  
SQ Sequence 563 AA;

Query Match 99.9%; Score 2999; DB 3; Length 563;  
Best Local Similarity 100.0%; Pred. No. 1e-236;  
Matches 563; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 MVRPQRMKKRFFILILNGVWNLVETSKDRPIAVPGSWNEQYODLCYEEGPFTYKTFYVVK 60
DB 1 MVRPQRMKKRFFILILNGVWNLVETSKDRPIAVPGSWNEQYODLCYEEGPFTYKTFYVVK 60
QY 61 XLSQKHIRLYFAAVNTDCEVFLNGEKGEMHIEYLPFEVDVTGKVGSGENELRVVENRL 120
DB 61 XLSQKHIRLYFAAVNTDCEVFLNGEKGEMHIEYLPFEVDVTGKVGSGENELRVVENRL 120
QY 121 KVGGFPSKVPDSTHTVGFPGSFPFANFDPFPGGIRPVLIETDHAIRLIDWDTSES 180
DB 121 KVGGFPSKVPDSTHTVGFPGSFPFANFDPFPGGIRPVLIETDHAIRLIDWDTSES 180
QY 181 EPEKLGKVKVIEVSEEAQEMTIKLGEEKKIRTSNRFVEGEFILENARFWSLEDPY 240
DB 181 EPEKLGKVKVIEVSEEAQEMTIKLGEEKKIRTSNRFVEGEFILENARFWSLEDPY 240
QY 241 LYPLKVELEKDEYTLDIGRTISWDEKRLYLNGKPVFLKGFGKHEEPVYGQGTFFPLMI 300
DB 241 LYPLKVELEKDEYTLDIGRTISWDEKRLYLNGKPVFLKGFGKHEEPVYGQGTFFPLMI 300
QY 301 KDFNLKWINANSFRTSHYPIYSEEWLDLADRLGILVIDEAPHVGITRYHNPETOKIAED 360
DB 301 KDFNLKWINANSFRTSHYPIYSEEWLDLADRLGILVIDEAPHVGITRYHNPETOKIAED 360
QY 361 NIRRMIDRHKNHPSVIMSVANEPESNHDAGEFFKALYETANEMDRTRPVVMVSMMDAP 420
DB 361 NIRRMIDRHKNHPSVIMSVANEPESNHDAGEFFKALYETANEMDRTRPVVMVSMMDAP 420
QY 421 DERTRDVALKYFDIVCVARYGYIYQGRIEBGLQALEKDI EELYARHKKPIFVTEGAD 480
DB 421 DERTRDVALKYFDIVCVARYGYIYQGRIEBGLQALEKDI EELYARHKKPIFVTEGAD 480
QY 481 AIAGIHYDPQMFSEYQALVEKTRILLKKDYIIGTHVAFADFKTPQNVRRPILNHK 540
DB 481 AIAGIHYDPQMFSEYQALVEKTRILLKKDYIIGTHVAFADFKTPQNVRRPILNHK 540
QY 541 GVFTTRDQPKLVAVHLRRLMSEV 563
DB 541 GVFTTRDQPKLVAVHLRRLMSEV 563
```

RESULT 2

AAW93820

ID AAW93820 standard; protein; 602 AA.

XX AAW93820;

AC AAW93820;

DT 25-JUN-1999 (first entry)

DE Bacillus sp. GUS protein.

XX

KW GUS; beta-glucuronidase; secreted; reporter molecule; marker;

KM receptor molecule; diagnostic tool; transgene construction; plant;

XX insect; cleavage; detoxification; glucuronide.

OS Bacillus sp.

XX

PN WO9913085-A2.  
XX  
XX PD 18-MAR-1999.  
XX  
XX PF 09-SEP-1998; 98WO-US019217.  
XX PR 09-SEP-1997; 97US-0058263P.  
XX PA (CAMP-) CAMBIA BIOSYSTEMS LLC.  
XX PI Jefferson RA, Kilian A, Keese PR;  
XX WPI; 1999-229241/19.

New isolated microbial beta-glucuronidase.

Claim 4; Fig 3; 76pp; English.

This invention describes a novel secreted form of *Bacillus* sp. beta-glucuronidase (BogUS). The microbial BogUS polypeptide can be used as a reporter/effector molecule and as a diagnostic tool. The products of the invention can be used as markers for transgene constructions, e.g. in plants or insects. They can also be used for the cleavage and detoxification of glucuronides and to examine conjugation patterns of glucuronides

SQ Sequence 602 AA;

Query Match 31.1%; Score 932; DB 2; Length 602;  
Best Local Similarity 35.6%; Pred. No. 2.7e-67;  
Matches 221; Conservative 94; Mismatches 223; Indels 82; Gaps 15;

```
QY 1 MVRPQRMKKRFFILILNGVWNLVETSKDRPIAVPGSWNE--QYODLC 44
DB 1 MVRPQRMKKRFFILILNGVWNLVETSKDRPIAVPGSWNE--QYODLC 44
QY 45 YEEGPFTYKTFYVVKXLSQKHIRLYFAAVNTDCEVFLNGEKGEMHIEYLPFEVDVTGK 104
DB 45 YEEGPFTYKTFYVVKXLSQKHIRLYFAAVNTDCEVFLNGEKGEMHIEYLPFEVDVTGK 104
QY 61 NHIGVWYEREFVTPAYLKDQRIVLRFGSATKAIYVNGELVVEHKGFLPEAEINNS 120
DB 61 NHIGVWYEREFVTPAYLKDQRIVLRFGSATKAIYVNGELVVEHKGFLPEAEINNS 120
QY 105 VSGENELRVVENRLKVGGFPSKVPDSTHTVGFPGS-----FPANFDPFPGY 154
DB 105 VSGENELRVVENRLKVGGFPSKVPDSTHTVGFPGS-----FPANFDPFPGY 154
QY 121 LNDGMNRVTVAVDNIL-----DSTLPVGLYSEHHEGLGKVIIRKNPNDFFENYA 170
DB 121 LNDGMNRVTVAVDNIL-----DSTLPVGLYSEHHEGLGKVIIRKNPNDFFENYA 170
QY 155 GIIRPVLIETDHAIRLIDWDTSESEPEKLGKVKVIEVSEEAQEMTIKLG--EE 212
DB 155 GIIRPVLIETDHAIRLIDWDTSESEPEKLGKVKVIEVSEEAQEMTIKLG--EE 212
QY 171 GHRFPVKIYTPPTFYVEDISVTDENGPF--TGTVTYTVDPQ---GKAETVKVSVDEE 223
DB 171 GHRFPVKIYTPPTFYVEDISVTDENGPF--TGTVTYTVDPQ---GKAETVKVSVDEE 223
QY 213 KKIRTSNRFVEGEFILENARFWSLEDPYLYPLKVELEKDEYTLDI-----GIRTSWDEK 267
DB 213 KKIRTSNRFVEGEFILENARFWSLEDPYLYPLKVELEKDEYTLDI-----GIRTSWDEK 267
QY 224 GKVVASTEGLSGNVEIPNVILMEPLNTLYQIKVELVNDGLTIDVYEEPFGRVTEBVNDG 283
DB 224 GKVVASTEGLSGNVEIPNVILMEPLNTLYQIKVELVNDGLTIDVYEEPFGRVTEBVNDG 283
QY 268 RLYLNGKPVFLKGFGKHEEPVYGQGTFFPLMIKDFNLKWINANSFRTSHYPIYSEEWLD 327
DB 268 RLYLNGKPVFLKGFGKHEEPVYGQGTFFPLMIKDFNLKWINANSFRTSHYPIYSEEWLD 327
QY 284 KFLINNKPFYFKFGKHEDTPINGRGFNEASNMDFNLIKWIGANSFRTAHYPIYSEELMR 343
DB 284 KFLINNKPFYFKFGKHEDTPINGRGFNEASNMDFNLIKWIGANSFRTAHYPIYSEELMR 343
QY 328 LADRLGILVIDEAPHVIGITRYHYN-----PETOKIA-----EDNIRRMIDR 368
DB 328 LADRLGILVIDEAPHVIGITRYHYN-----PETOKIA-----EDNIRRMIDR 368
QY 344 LADREGLVVIDETPAVG---HLNFMATGGLGEGSERVSTWEKIRTFEHHQDVLRELVS 400
DB 344 LADREGLVVIDETPAVG---HLNFMATGGLGEGSERVSTWEKIRTFEHHQDVLRELVS 400
QY 369 HKNHPSVIMSVANEPESNHDAGEFFKALYETANEMD--RTRPVVMVSMMDAPDERTRDV 427
DB 369 HKNHPSVIMSVANEPESNHDAGEFFKALYETANEMD--RTRPVVMVSMMDAPDERTRDV 427
QY 401 DKNHPSVIMSVANEPESNHDAGEFFKALYETANEMD--RTRPVVMVSMMDAPDERTRDV 458
DB 401 DKNHPSVIMSVANEPESNHDAGEFFKALYETANEMD--RTRPVVMVSMMDAPDERTRDV 458
QY 428 ALKYFDIVCVARYGYIYQGRIEBGLQALEKDI EELYARHKKPIFVTEGADATA 483
DB 428 ALKYFDIVCVARYGYIYQGRIEBGLQALEKDI EELYARHKKPIFVTEGADATA 483
QY 459 VAEILIDVIALNRNGWYFDGDDLE---AAKVHRLQEFHAWNKRCPGKPIIMITEYGADTVA 515
DB 459 VAEILIDVIALNRNGWYFDGDDLE---AAKVHRLQEFHAWNKRCPGKPIIMITEYGADTVA 515
QY 484 GIHYDPQMFSEYQALVEKTRILLKKDYIIGTHVAFADFKTPQNVRRPILNHKGV 543
DB 484 GIHYDPQMFSEYQALVEKTRILLKKDYIIGTHVAFADFKTPQNVRRPILNHKGV 543
QY 516 GFHDIDPVWTEEYQVEYYQANHVVFDEFENFVGEQANMFADFATSQGVNRVQGNKKGVF 575
DB 516 GFHDIDPVWTEEYQVEYYQANHVVFDEFENFVGEQANMFADFATSQGVNRVQGNKKGVF 575
QY 544 TRDRQPKLVAVHLRRLMSEV 563
```

Db 576 TRDRKPKLAHVFRERWTNI 595

RESULT 3  
AAW93825  
ID AAW93825 standard; protein; 602 AA.

XX AAW93825;  
DT 25-JUN-1999 (first entry)

XX Bacillus sp. codon optimised GUS protein.

XX GUS; BogUS; beta-glucuronidase; secreted; reporter molecule; marker;  
KW receptor molecule; diagnostic tool; transgene construction; plant;  
KW insect; cleavage; detoxification; glucuronide.

XX Bacillus sp.

PN WO9913085-A2.

PD 18-MAR-1999.

PF 09-SEP-1998; 98WO-US019217.

PR 09-SEP-1997; 97US-0058263P.

PA (CAMB-) CAMBIA BIOSYSTEMS LLC.

PI Jefferson RA, Killian A, Keese PK;

DR WPI; 1999-229241/19.

DR N-PSDB; AAX23826.

PT New isolated microbial beta-glucuronidase.

XX Example 3; Fig 13A-D; 76pp; English.

CC This invention describes a novel secreted form of Bacillus sp. beta-  
CC glucuronidase (BogUS). The microbial BogUS polypeptide can be used as a  
CC reporter/effector molecule and as a diagnostic tool. The products of the  
CC invention can be used as markers for transgene constructions, e.g. in  
CC plants or insects. They can also be used for the cleavage and  
CC detoxification of glucuronides and to examine conjugation patterns of  
CC glucuronides

XX Sequence 602 AA;

Query Match 31.0%; Score 929; DB 2; Length 602;  
Best Local Similarity 35.6%; Pred. No. 4.8e-67;  
Matches 221; Conservative 93; Mismatches 224; Indels 82; Gaps 15;

OY 1 MVRPQRNKKRIFILINGVWNLEV-----TSKDRPIAVPGSWNE--QYQDLC 44

DB 1 MLYPINTETRGVFDLNGVMNFKLDYGKLEKMYESKLTDTISMAVPSYNDIGVTKAIR 60

OY 45 YEEGPFTYKTFYVVPKXLSQKHRLYFAAVNTDCEVFLNGEKVGENHIEYLPEVDVTGK 104

DB 61 NHIGYVMYEREFVPAVLKDDQLVLRFGSATHKAIVYVNGELVVEHKGFLEPFEAEINNS 120

OY 105 VKSGENELRVVVENRLKVGSPSKVPDSGTHVGFSGS-----FPPANFDFFPYG 154

DB 121 LRDMNRVTAVDNIL-----DDSTLPVGLYSERHHEGLGKVIIRKNKPNDFFNVA 170

OY 155 GIIRPVLIIEFTDHARILDIWVDTSESEPEKKLGKVKIIEVSEBAVGQEMTIKG--EEB 212

DB 171 GLHRPVKIYTFPFTYVEDISVTDENGPR--TGTVTYVDVFG---GKAETVKVSVVDEE 223

OY 213 KKIRTSNRFVEGEFILLENAFNSLEDPLYLPLKVELEKDEYTLDI-----GIRTI SMDEK 267

DB 224 GKVVASTEGLSGNVEIPNVILMEPLNTYLYQIKVELVNDGLTIDVYBEFPVGRVTEVNDG 283

OY 268 RLYLNGKPVFLKGFGEKHEEPVLGQGTFFYPLMIKDENLKWINANSFRTSHYPYSEWLD 327

DB 284 KEFINNKPFFYFKGFGKHEDTPINGRGFNASNVMDFNLIKWIGANSFRTAHYPYSEELMR 343

OY 328 LADRLGILVIDEAPHVGITRHYHN-----PETOKIA-----EDNIRRMIDR 368

DB 344 LADREGLVVIDETPAVG--HINFMATTGLGEGSERVSTWEKIRTFEHHQDVLRELVS 400

OY 369 HKNHPSVIMSVANEPESNHDPAGEFFKALYETANEMD-RTRPVVMVSMDBPDERTRDV 427

DB 401 DKNHPSVVMWSIANEAATEEGAYEYFKPLVELTKELDPQKRPVTIVLFVMAIPE--TDK 458

OY 428 ALKYFDIVCVNRYGYIYQGRIEGLQALEKDIIELYARHR---KPIFTEFGADAIA 483

DB 459 VAEILIDVIALNRYNGWYFDGDLLEAAKVHLR---QEFHAMNKRCPGKPIMITFEGADTVA 515

OY 484 GIHYDPQMFSEBYQAEIYEKTI RLLKKDVIIGTHVMAFADFKTPQNVRRPILNHKGVF 543

DB 516 GFHDIDPVMFTEBYQVEYQANHVVFEDEFENVEGEQAMNFADFATSQGVMRVQGNKKGVF 575

OY 544 TRDRQPKLVAVHLRLWSEV 563

DB 576 TRDRKPKLAHVFRERWTNI 595

RESULT 4  
AAW93822  
ID AAW93822 standard; protein; 602 AA.

XX AAW93822;

DT 25-JUN-1999 (first entry)

XX Bacillus sp. GUS protein.

KW GUS; BogUS; beta-glucuronidase; secreted; reporter molecule; marker;  
KW receptor molecule; diagnostic tool; transgene construction; plant;  
KW insect; cleavage; detoxification; glucuronide.

XX Bacillus sp.

PN WO9913085-A2.

PD 18-MAR-1999.

PF 09-SEP-1998; 98WO-US019217.

PR 09-SEP-1997; 97US-0058263P.

PA (CAMB-) CAMBIA BIOSYSTEMS LLC.

PI Jefferson RA, Killian A, Keese PK;

DR WPI; 1999-229241/19.

PT New isolated microbial beta-glucuronidase.

XX Example 1; Fig 5; 76pp; English.

CC This invention describes a novel secreted form of Bacillus sp. beta-  
CC glucuronidase (BogUS). The microbial BogUS polypeptide can be used as a  
CC reporter/effector molecule and as a diagnostic tool. The products of the  
CC invention can be used as markers for transgene constructions, e.g. in  
CC plants or insects. They can also be used for the cleavage and  
CC detoxification of glucuronides and to examine conjugation patterns of  
CC glucuronides

XX Sequence 602 AA;

Query Match 31.0%; Score 929; DB 2; Length 602;  
Best Local Similarity 35.6%; Pred. No. 4.8e-67;

Matches 221; Conservative 93; Mismatches 224; Indels 82; Gaps 15;

```

Qy      1 MVRPQNRKKRFILLINGVWNLEV-----TSKDRPIAVPGSMNE--QYODLC 44
Db      1 MLXPINTETRGVFDLNGVWNFKLDYGKGLUEKEWKYESKLTDTISMAVPSSYNDIGVTKEIR 60
Qy      45 YEEGPFYTKTFEYVPKLSQKHIRLYFAVNTDCEVFLNGEKGVENHIEYLPEFVDVTGK 104
Db      61 NHIGVWYEREFETVPAYLKQORIVLRFGSATHKAIVYVNGELVVEHHKGFELPEAEINNS 120
Qy      105 VKSGENELRVVVENRLKVGCFPSKVPDSGTHTVGFEGS-----FPANFDFFPYG 154
Db      121 LRQGMNRVTVAVDNL-----DDSTLPVGLYSEHHEGLGKVIKKNKPNDFPFNYA 170
Qy      155 GTRPVLIEFTDHARLIDIVWDTSESEBEKKLGKVKVKEIVSEAVGQEMITKLG--EEE 212
Db      171 GLHRPVKIYTPPTFYVEDISVTVDFNGP--TGTVTVYVDFQ---GKAETVKVSVDEE 223
Qy      213 KKRITSNRFEVEGEFIENARFWSLEDPLIPLKVELKDEYTLDI---GIRTISWDEK 267
Db      224 GKVAVASTEGLSGNVEIPNVILMEPLNTLYLQIKVELVNDGLTIDVYEEFPGRVTEVANDG 283
Qy      268 RLYNGKPVFLKGFSGHHEEPVLGGQTFPYPLMIKDPNLKKMINANSFRSHYPYSEEWLD 327
Db      284 KFLINNKPFYFKGFGKHEDTPINGRGFNEASNVMDFNLKMGIGANSFRTAHYPSSEELMR 343
Qy      328 LADRLGILVIDEAPHVGITRYHYN-----PETOKIA-----EDNIRRMIDR 368
Db      344 LADREGLVIDETPAVGV--HNFMAATGLGEGSERVSTWEKIRTFEHHQDVLRELVS 400
Qy      369 HKNPSVIMMSVANEPESNHPDAEGFFKALYETANEMD-RTRPVVMVSMMDAPDERTRV 427
Db      401 DKMHPSVMMMSIANEAATEEBGAYEYFRPLVELTKELDPQKRPVITLVEFMATPE--TDK 458
Qy      428 ALKXFVIVCVNRRYVGWYIYQGRIEEGLQALEKDIIEELYARHR---KPIFVTEFGADATA 483
Db      459 VAEILDVIALNRYNGWYFDGGLDEAAKVHLR---QEFHANMKRCPGKPRIMITEYGADTVA 515
Qy      484 GIHYDPQPMFSEEYOAELEVETIRLLKKDYIIGTHVAFADFKTPONVRRPILNHHGVF 543
Db      516 GFHDIDPVMFTEEYQVEYYQANHVVFEDEFENFVGEGQAMNPFADFATSQGWRVQGNKKGVF 575
Qy      544 TRDRQPKLVAAHLRLMSEV 563
Db      576 TRDRKKPKLAAHVFRERWTNI 595

```

XX		
PI	Jefferson RA,	Killian A, Keese PK;
XX		
PA	(CAMB-) CAMBIA BIOSYSTEMS LLC.	
PR	09-SEP-1997,	97US-0058263P.
XX		
PF	09-SEP-1998,	98WO-US019217.
XX		
PD	18-MAR-1999.	
XX		
PN	WO9913085-A2.	
OS	Bacillus sp.	
XX		
KW	GUS; BoGUS; beta-glucuronidase; secreted; reporter molecule; marker;	
KM	receptor molecule; diagnostic tool; transgene construction; plant;	
XX	insect; cleavage; detoxification; glucuronide.	
DE	Bacillus sp. GUS protein fragment.	
DT	25-JUN-1999	(first entry)
AC	AAW93826;	
ID	AAW93826 standard; protein; 602 AA.	
RESULT 5		

DR	WPI; 1999-229241/19.
XX	
PT	New isolated microbial beta-glucuronidase.
XX	
PS	Example 3; Fig 13D; 76pp; English.
XX	
CC	This invention describes a novel secreted form of <i>Bacillus</i> sp. beta-glucuronidase (Bogus). The microbial Bogus polypeptide can be used as
CC	reporter/effector molecule and as a diagnostic tool. The products of
CC	invention can be used as markers for transgene constructions, e.g. in
CC	plants or insects. They can also be used for the cleavage and
CC	detoxification of glucuronides and to examine conjugation patterns of
CC	glucuronides
XX	
XX	
5Q	Sequence 602 AA;

Query Match	31.0%;	Score 929;	DB 2;	Length 602;
Best Local Similarity	35.6%;	Pred. No. 4.8e-67;		
Matches 221; Conservative	93;	Mismatches 224;	Indels 82;	Gaps 15;

[illegible]

RESULT 6	
AAB28402	
ID	AAB28402 standard; protein; 602 AA.
XX	
AC	AAB28402;
XX	
.DT	26-JAN-2001 (first entry)
XX	



Dd	224	GKVAASTEGLSGNVEIPNVILWEPLNTLYLOIKVELVNDGLTIDVYEEBFGVRTVEAVNDG	283
Qy	268	RLYNGKPVFLKFGKGHEEPVLGGTFYPLMKDFNLKKWINANSFRSHYPYSEEWLD	327
Dd	284	KPLINNKPFYFKGFGKHEDTPINGRGFNEASNVMDFNILKMIGANSPRTAHYPYSEELMR	343
Qy	328	LADRLGILVIDEAPHVGITRYHYN-----PETOKIA-----EDNIRRMIDR	368
Dd	344	LADREGLVIVIDETPAVG--HLNFMAATTGLGEGSERVSTWEEKIRTEHHQDVLRELVSr	400
Qy	369	HKNPSSVIMSVANEPEESNHDPDAEGFPKALYETANEND-RTRPVVWSMMDAPDERTRDV	427
Dd	401	DKNPSVMMWSIANEAATEEBGAYEFKPLVELTKELDPOKRPTVILVFWMATPE--TDK	458
Qy	428	ALKYFDIVCVNRYYGYIYOGRIBESLQALEKDIEELYARHR---KPIFVTEFGADAIA	483
Dd	459	VAEILDVIALNRRYNGWYFDGGDLBAKVHLR--QEPhAMNKRCPGKPIMITeYGADIVA	515
Qy	484	GIHYDPOMSEBYOAELEVektirLLKKDYIIgTHWAfADFKTPONVRrpILNHKGvF	543
Dd	516	GFHDIDPMFTEEYOVEYYQANHVVDEFENFVGQAMNFADFATSOGMRVOGNKKGVF	575
Qy	544	TRDRQPKLVAAHVLRRLMWSEV	563
Dd	576	TRDRKPKLAHVFRERWTNI	595

RESULT 7  
 AAW93821  
 ID AAW93821 standard; protein; 618 AA.  
 AC AAW93821;  
 DT 25-JUN-1999 (first entry)  
 DE Bacillus sp. GUS protein.  
 KW GUS; BoGUS; beta-glucuronidase; secreted; reporter molecule; marker;  
 KW receptor molecule; diagnostic tool; transgene construction; plant;  
 KW insect; cleavage; detoxification; glucuronide.  
 OS Bacillus sp.  
 PN WO9913085-A2.  
 PD 18-MAR-1999.  
 PF 09-SEP-1998; 98WO-US019217.  
 PR 09-SEP-1997; 97US-0058263P.  
 PA (CAMB-) CAMBIA BIOSYSTEMS LLC.  
 PI Jefferson RA, Kilian A, Keese PK;  
 DR WPI; 1999-229241/19.  
 DR N-PSDB; AAX23825.  
 PT New isolated microbial beta-glucuronidase.  
 PS Claim 5; Fig 4A-C; 76pp; English.  
 CC This invention describes a novel secreted form of Bacillus sp. beta-  
 CC glucuronidase (BoGUS). The microbial BoGUS polypeptide can be used as a  
 CC reporter/effector molecule and as a diagnostic tool. The products of the  
 CC invention can be used as markers for transgene constructions, e.g. in  
 CC plants or insects. They can also be used for the cleavage and  
 CC detoxification of glucuronides and to examine conjugation patterns of  
 CC glucuronides  
 SQ Sequence 618 AA;



Db 467 LIDVIALNRVNGWYFDGDLBAKVHLR---QEFHAMNKRCPGKPIMITETYGADTVAGFH 523  
QY 487 YDPQPMFSEEXQAEIVKTIIRLLKKDYIIGTHVWAFADFKTPQNVRRPILNHKGVFTRD 546  
Db 524 DIDPVMFTBEEYQVEYYQANHVVDEFENFVGEQAMNFADFATSGVMRVQGNKKGVFTRD 583  
QY 547 RQPKLVAVHLRLRMSEV 563  
Db 584 RKPKLAAHVFRERWTNI 600

RESULT 9  
AAB28409  
ID AAB28409 standard; protein; 602 AA.

XX AAB28409;  
AC  
XX  
DT 26-JAN-2001 (first entry)  
XX  
De Salmonella beta-glucuronidase.  
XX  
KW Microbial; beta-glucuronidase; GUS; Enterobacter; Salmonella;  
KW Pseudomonas; Staphylococcus; Thermotoga; transgenic plant; bioindicator;  
KW transgenic insect; marker; glucuronide detoxification.  
XX  
OS Salmonella sp.  
XX  
PN WO200055333-A1.  
XX  
PD 21-SEP-2000.  
XX  
PF 16-MAR-2000; 2000WO-US007107.  
XX  
PR 17-MAR-1999; 99US-00270957.  
XX  
PA (CAMB-) CAMBIA BIOSYSTEMS LLC.  
XX  
PI Jefferson RA, Mayer JE;  
XX  
DR WPI; 2000-647075/62.  
DR N-PSDB; AAA07939.  
XX  
PT Novel microbial beta-glucuronidase genes and gene products used as  
PT reporter/effector molecule, as diagnostic tool, in positive selection, to  
PT target molecules to specific cells and to detect and track linked genes.  
XX  
PS Example 3; Fig 17; 116pp; English.

XX The present sequence is a microbial beta-glucuronidase (GUS) protein. GUS  
CC genes were obtained from six different genera: Enterobacter/Salmonella,  
CC Pseudomonas, Salmonella, Staphylococcus and Thermotoga. Microbial GUS can  
CC be used as a reporter/effector molecule for transgenic constructions and  
CC in in vitro diagnostic applications. It may also be used to generate  
CC sentinel plants that serve as bioindicators of environmental status. It  
CC may be used to generate transgenic insects for tracking insect  
CC populations or to facilitate the development of a bioassay for compounds  
CC that affect molecules critical for insect development (e.g. juvenile  
CC hormone). Secreted GUS may also serve as a marker for beneficial fungi  
CC destined for release into the environment. In animal systems, secreted  
CC GUS may be used to achieve extracellular detoxification of glucuronides  
CC (e.g. toxin glucuronide) and to examine conjugation patterns of  
CC glucuronides. Microbial GUS may also be used in traditional medical  
CC diagnostic assays, for drug testing, pharmacokinetic studies,  
CC bioavailability studies, diagnosis of diseases and syndromes, following  
CC progression of disease or its response to therapy. Microbial GUS has  
CC increased thermal stability, high turnover number and enzymatic activity.  
CC It is highly specific for the substrate and water soluble, and the  
XX substrates are stable  
SQ Sequence 602 AA;

Query Match 30.7%; Score 920; DB 3; Length 602;

Best Local Similarity 35.3%; Pred. No. 2.6e-66;  
Matches 213; Conservative 110; Mismatches 223; Indels 58; Gaps 11;

QY 1 MVRPQRNKKRPFILINGVNMLEVTSKD-----RPIAVPGSWNEQY--QDL 43  
Db 1 MLRSVETATAEIKKLDGLMFCMDSEECGNAQQWWRQPLPQSRALIAVPGSYNDQFAAAEI 60  
QY 44 CYEEGPFYTKTFYVPPKXLSQKHIRLYFAAVNTDCEVFLNGEKVGENHIEYLPFEVDVTG 103  
Db 61 RNYYGNVWVYQREIRIPKGMQRQRIVLRFDAVTHYQKVVNDQFLMEHQGYTPFEADISH 120  
QY 104 KYKSGEN-ELRVVENRLKVGGFPSKVPDGGTHTVGFFGSPFPANF-DFPYGGLIRPVL 161  
Db 121 LISAGESVRITVCVNNELMNGTIP-----PGVVTQGVNGKKQQAIFYHDFNYAGIHR SVM 175  
QY 162 IEFTDHARILDIWVDTSSESEPEKKLGKVKYKIEVSEAVGQEMTIKLGEEKKIRTSNRF 221  
Db 176 LYTTPKTFVEDITVVTQVAD---DLAQATVAMQVRANG---EVRVELRDAEQQLVASGQG 229  
QY 222 VEGFIFLENARFWSLEDPIYPLKV----ELEKDEYTLDIGRTISWDEKRLYNGKPVF 277  
Db 230 EKGELLLEGPRLMQPEGVYELRVIAQHODEODEYPLRVGIRSVVEVKGEQPLINHKPFY 289  
QY 278 LKGFQKHEEPFVLGQGTFFYPLMIKDFNLKWINANSFRTSHYPYSEEWLDLDRGLIVT 337  
Db 290 FTGFRHEDADLRGKGFQDNVLMVHDHALMWDIGANSYRTSHYPAEEMLDMADEHGVIT 349  
QY 338 DEAPHVGI-----TRYHNPETOKIAEDNIRRMIDRHQNPVSIMWS 379  
Db 350 DETAAVGFNLSLGISFDVGEKPKELYSDEAVNDETQRAHLQAIKELIARDXNPVSVMWS 409  
QY 380 VANEPSNHPDAEGFFKALYETANEMDRTPVVMVSM--DAPDERTRDVALKYFDIVCV 437  
Db 410 IANEPDTRPNGAREYFAPLAQATRELDPTRPITCVNMFCDAESDTITDL----FDVCL 465  
QY 438 NRYYGWYIYQGRIEEGLQALEKDIEELYARHRKPIFVTEFGADAIAGIHYPQMFSEY 497  
Db 466 NRYYGWYVQGDLEKAEKYLEKELLAMQEKLHRPIITTEYGVDTLAGLSMYNDMSEY 525  
QY 498 QAEIVKTIIRLLKKDYIIGTHVWAFADFKTPQNVRRPILNHKGVFTRDQPKLVAVHLR 557  
Db 526 QCAMLDMYHRVFDVSAAVGEQVWNFADFATSGIMRVGKNKGIFTRDRKPSAAFLQ 585  
QY 558 RLWS 561  
Db 586 KRW 589

RESULT 10  
AAM48998  
ID AAM48998 standard; protein; 598 AA.  
XX  
AC AAM48998;  
XX  
DT 10-MAY-2002 (first entry)  
XX  
DE L gasseri beta-glucuronidase.  
XX  
KW Beta-glucuronidase; GUS; low pH; Lactobacillus.  
XX  
OS Lactobacillus gasseri.  
XX  
PN WO200190305-A2.  
XX  
PD 29-NOV-2001.  
XX  
PF 22-MAY-2001; 2001WO-US016667.  
XX  
PR 23-MAY-2000; 2000US-0206372P.  
XX  
PA (UYNC-) UNIV NORTH CAROLINA STATE.  
XX  
PI Russell WM, Klaenhammer TR;

XX WPI: 2002-062529/08.  
DR N-PSDB; AAL44836.  
XX  
PT Novel isolated nucleic acid of Lactobacillus gasseri encoding beta  
PT glucuronidase having activity at acidic pH which is useful as reporter  
PT protein for highly aciduric organisms, or as marker of transformed cells.  
XX  
XX  
PS Claim 19; Page 49-51; 53pp; English.  
XX  
XX The present invention provides the protein and coding sequences of  
CC Lactobacillus gasseri beta-glucuronidase (GUS). The protein has maximum  
CC activity at acidic pHs, and thus can be used to detect low pH  
CC environments. The present sequence is the protein of the invention  
XX  
SQ Sequence 598 AA;

Query Match 30.6%; Score 919.5; DB 5; Length 598;  
Best Local Similarity 34.1%; Pred. No. 2.9e-66;  
Matches 207; Conservative 110; Mismatches 227; Indels 63; Gaps 14

[illegible]

RESULT	11
ID	ADF28924
XX	ADF28924 standard; protein; 648 AA
AC	ADF28924;
XX	
DT	12-FEB-2004 (first entry)

XX		
DE	Murine beta-glucuronidase precursor.	
XX		
KW	Mouse; beta-glucuronidase; gene therapy; cytostatic; antidiabetic;	
KW	immunosuppressive; hepatotropic; adeno-associated virus.	
XX		
OS	Mus sp.	
XX		
FH	Key	Location/Qualifiers
FT	Peptide	1..22
FT		/label= Signal_peptide
FT	Protein	23..648
FT		/label= Beta-glucuronidase

PN WO2003089011-A1.  
XX  
XX 30-OCT-2003.  
PD  
XX  
XX 21-APR-2003; 2003WO-US012324.  
PF  
XX 19-APR-2002; 2002US-0374083P.  
PR

PA (UYFL ) UNIV FLORIDA.  
XX  
PI Atkinson MA, Flotte TR, Song S, Lohler SA,  
XX  
DR WPI; 2003-845502/78.

**PT** New adeno-associated viral vector, useful in preparing a composition for PT treating or preventing e.g., cancer, diabetes, or autoimmune, pancreatic PT or liver disease.

Example 4; SEQ ID NO 23; 183pp; English.

The present sequence is that of murine beta-glucuronidase precursor (beta-D-glucuronoside glucuronohydrolase). This is an example of a therapeutic protein that can be encoded by an adeno-associated virus (AAV) vector of the invention. Such vectors comprise a promoter operably positioned upstream of a nucleic acid encoding a biologically-active therapeutic mammalian serpin or cytokine polypeptide, and optionally also include an enhancer sequence and a post-transcriptional regulatory sequence. A recombinant AAV virion comprising the vector, and a mammalian cell (preferably an endothelial, islet, hepatocyte, pancreas, kidney, muscle, spleen, liver, heart, lung, or brain cell) comprising the vector are claimed. A claimed composition comprises the vector, the recombinant AAV virion, AAV viral particles, or the mammalian cell, and is used in cancer, diabetes, autoimmune disease, pancreatic disease or liver disease therapy. The composition is also used in claimed methods for preventing type I diabetes, and for reducing the rate of disease progression of type I diabetes, in a human.

Sequence 648 AA;

Query Match	30.1%;	Score 902.5;	DB 7;	Length 648;
Best Local Similarity	35.7%;	Pred. No. 7.9e-65;		
Matches 223; Conservative	103;	Mismatches 207;	Indels 91;	Gaps 21;

[illegible]

Qy	202	QEMTITKLGEEEKKIRTSNRFVEGEFIELENARFW----	SLEDP-YYLPKY-----ELEKD	251
Dd	256	-OLEVQLDDEDGKVVAHGTCNOGOLQVPYSANLWMPYLMHEHPAYMYSLEVKVTTTESVTD		314
Qy	252	EYTLDIGIRTISWDEKRLYLNGKPVFELKGFGKHEEFPVLGOGTFFPLMKDPENLLKWINA	311	
Dd	315	YTLLPVGIRTVAVTSKFLINGKPFYPQGVNKHEDSDLRGKGFDMPLLVKDFNLRLWLGA	374	
Qy	312	NSFRTSHYPYSEEMDLADRLGLVLIDEAPHVGIT-----RYHNPETOKIAEDN	361	
Dd	375	NSFRTSHYPYSEEVLOLCDRYGIVVIDECRGVIGI.VLPGSGNESLRRHL-----EVMEEEL	429	
Qy	362	IRRMIDRHKNHPSVIMSVANEPESNHDAEGFFKALYETANEMDRTPVNVMS--NMMA	419	
Dd	430	VR----RDKNHPAUVWMSVANEPSSALKPAAYYFKLT.LTHTKALDLTRPYTFVSNAKYDA	485	
Qy	420	PDERTRDVAl.KYFDIVCVNRYYYGWYIYOGRIEBGLQAULEKDI.EEL.YARRHKPI.FVTEFGA	479	
Dd	486	-----DLGARVVDDVICVNSYFSWMHYGHLEVIQPQLNSQFENWYKTHQKPIIOSEYGA	539	
Qy	480	DALAGIHYPD.PQMFSSEYQOAE.LVEKTIRLL--LKDYILIGTHWAFADEFKTPQNVRPRIL	537	
Dd	540	DALPGIHEDPPRMFSESEYQKAVLENHVS.LDQRKEAYVGEL.IWNFADEMTNQSP.LRVIG	599	
Qy	538	NHKGVFTRDRQPKLVAAHVL-R.LW 560		
Dd	600	NKKGI.FTRORQPKTSAFILRERYW 623		

RESULT 12  
AAE02444  
ID AAE02444 standard; protein; 648 AA.

DT	10-AUG-2001	(first entry)
XX		
DE		
XX		
XX		
KW	Murine, adeno-associated viral expression vector; AAV; gene therapy;	
KW	lysosomal storage disease; LSD; mucopolysaccharidoses VII; MPS VII;	
KW	Sly syndrome; beta-glucuronidase; GUS; glycosaminoglycan; GAG.	
XX		
OS	Mus musculus.	

Key	Location/Qualifiers
Peptide	1..26
FT	/label= Signal_peptide
FT	27..648
Protein	/note= "Murine mature beta-glucuronidase (GUS) "
FT	

PN WO200136603-A2.

PD 25-MAY-2001.

PF 17-NOV-2000; 2000WO-US031688.

PR 17-NOV-1999; 99US-0166097P.

PR 30-JUN-2000; 2000US-0215430P.

PA (AVIG-) AVIGEN INC.

PA (CHIL-) CHILDREN'S HOSPITAL MEDICAL CENT.

PI Podsakoff G, Watson G, Couto LB, Yang B;

XX

DR N-PSDB; AAD06387.

XX XX

PT protein defective or missing in lysosomal storage disease, in the

PT manufacture of a medicament for treating the lysosomal storage disease

PS Disclosure; Page 85-87; 97pp; English.

XX The present invention relates to recombinant adeno-associated virus (AAV)  
CC expression vectors and virions, which include genes coding for enzymes  
CC defective or missing in lysosomal storage disease (LSD). AAV is useful in  
CC the manufacture of a medicament for treating lysosomal storage disease  
CC e.g., mucopolysaccharidoses VII (MPS VII). MPS VII (Sly Syndrome) is due  
CC to the deficiency in the lysosomal enzyme beta-glucuronidase (GUS) which  
CC aids in the breakdown of glycosaminoglycans (GAGs). AAV is used in gene  
CC therapy. The present sequence is murine GUS protein. This sequence is  
CC used in AAV constructs  
XX  
SQ Sequence 648 AA;

Query Match	30.0%;	Score 899.5;	DB 4;	Length 648;
Best Local Similarity	35.9%;	Pred. No. 1.4e-64;		
Matches 224;	Conservative 101;	Mismatches 208;	Indels 91;	Gaps 21;

```

QY      1 MWPRQNRKKRFILLINGVNWLEV-TSKDR-----PIAVGSGWNQ 39
      27 | : : : : | : : : : | : : : : | : : : : | : : : :
Db      27 MLEPKESPSELKALDGLMHFRADISNNRLOGFEQGYRQPLRESGPVLDMVPSSFNDI 86

```

```

Oy      40  YODLCYEE--GPFYKKTFFYVPKXSQ---KHIRLYFAAVNTDCEVEFNKEKVGENHIEY  94
      | : | | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      87  TQEAALRDFIGWVYEREAIIIPRWTQDMDRVLRINSAHYYAAVWVNGIHVEHEGGH  146

```

```

Qy      95 LPEFVDVTGKVKSG---ENELRVVVENRLKVGFPSPKVP-----DSGTHTVGFEFGSF 143
      |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      147 LPEFADISKLVGSGPLTTCRITTAIINNTLT---PHTLPGTIVYKTDITSMYPKGYF--V 200

```

```

Qy      144  PEANFDPPPYGGIIRPVLIIEFTDHARILDIWDWDTSESEPEKLGKVK--VKIEVSEEAVG 201
          :||| | | :| :| | | | | :| :| | | :| :| |
Db      201  QDTSDFENYAGLHRSVLYTPTTYIIDITVITN--VEQDIGLVTYMISVGSEHF-- 255

```

```

QY 202 QEMTIKLGEBEKKIRTSNRFVEGEFIIENARFW----SLEDP-YLYPLKY-----ELEKD 251
      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 256 -QLEVQLDDEGSKYVAHGHTGNGQGLQVPSANLWMPYLMHNHPRAMYSLEVKVTTTESVTD 314

```

```
OY      252 EYTLDIGIRTISNDEKRLYLNGKRVFLKGFGKHFFPVLGQGTFFPLMIKD FNLLKWINA 311
```

|||||:: : : |||| : ||| : ||| ::||| ||| : |

```
Db      315 YTTLPIGIRTVATKSFLINGKPFYFQG VNKHESD IRGKFDPWPLLVKD FNLRLWLGA 374
```

```
QY      312 NSFRTSHYPYSEEWLDLADRLGLVIDEAPHVGIT-----RYHNPETQKIAEDN 361
          ||||| | | | | : | | | | | | | | | | | | | | | | | | | | | |
Db      375 NSFRTSHYPYSEEVQLCDRHYGIWIDECPGVGIILPQSGFNESLRHL-----EVMEEL 429
```

```

Qy      362  IRRMIDRHKNHPSVIMSVANEPSSNHPDAEGFFKALYETANEMDRTRPVVMVS--MMDA 419
      :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      430  VR----RDKNHPRVVMMSVANEPSSALKPRAYYFKLLIHTHKALDLTRPVTFVSNAKYDA 485

```

```
Oy      420 PDERTRDVALKKYFDIVCVNRRYYGWVIYQGRIEEGLALEKDIIEELVARRHKPIFVTEEGA 479
          | : ||| | : || : ||| | : ||| | : |||
Db      486 -----DLGARVYDVICVNSYSFWTHDYHLEVLQPOLNSQFENNMYTKHQKPRIQSEYGA 539
```

```
Qy      480 DAIAIHYPDPQMFSESEYQAEIVKTRLL--LKKDYIIIGTHVMAFADFKTPQNVRRPLL 537
      |||||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db      540 DAIPGIHEDPRMFSESEYQKAVLENYHSLVDQKKREYVVGELIIMFADFMTNQSPILRVIG 599
```

QY 538 NHKGVFTRDRQPKLVAVHLR-RLW 560  
| | : | | | | : | |  
Db 600 NKKGI FTRQRQPKTSAFILRRRYW 623

RESULT 13  
ABB70164  
ID ABB70164 standard; protein; 643 AA.

AC ABB70164;

XX

XX

DE Drosophila melanogaster SEQ ID NO 37284.

**KW** **Drosophila; developmental biology; cell signalling; insecticide;**



KW pharmaceutical.  
 XX  
 OS Drosophila melanogaster.  
 XX  
 PN WO200171042-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PF 23-MAR-2001; 2001WO-US009231.  
 XX  
 PR 23-MAR-2000; 2000US-0191637P.  
 PR 11-JUL-2000; 2000US-00614150.  
 XX  
 PA (PEKE ) PE CORP NY.  
 XX  
 PI Venter JC, Adams M, Li PWD, Myers EW;  
 XX  
 DR WPI; 2001-656860/75.  
 DR N-PSDB; ABL14267.  
 XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signaling and cell-cell  
 PT interactions.  
 XX  
 PS Disclosure; SEQ ID NO 37284; 21pp + Sequence listing; English.

CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from *Drosophila*. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABU16176-ABU30511); expressed DNA  
CC sequences (ABU01840-ABU16175) and the encoded proteins (ABB57737-  
CC ABB72072). The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://ftp.wipo.int/pub/published_pct_sequences)  
XX  
SQ Sequence 643 AA;

Query Match	29.9%;	Score 897;	DB 4;	Length 643;
Best Local Similarity	34.2%;	Pred. No. 2.2e-64;		
Matches 216;	Conservative 122;	Mismatches 205;	Indels 88;	Gaps 20;

```

QY 1 MVRPQNRKKRFFILILNGVNL-----EVTSKDR---PIAVGSMNEQ 39
   | : : : : | : : |
Db 1 MLYPRESETRVRSLDGIWNFVRSDQANPTQGVREDEWYAKELSKSRPTIPMPYVASYNDI 60
QY 40 YQD-LCYEEGPFYKTFYVPKLSQ-KHIRLYEAAVNTDCEVPLNGEKVGENHIEYLPF 97
   | | | | | | : : | : : | : : | : : | : : | : : |
Db 61 TTDNLRDHVGTWYDRKFFVPRSWSKDQRIWLRFGSVHYEAYVWINGQKVXKHEMGHLPF 120
QY 98 EVDVYTKVYKSG-ENELRVVVENRLKVGFP---SKVPDSGTHTVGFFGSPANFDFFP 152
   | : : : : | : : | : : | : : | : : | : : | : : |
Db 121 EAEVTDLISYGAENRITWCDNALIQTTVPQGRITEVENDGMTI-----VQSYTFDFFN 175
QY 153 YGGIRBPVLIETD HARILDIWBDTSESEPEKKLGKVKKIEVSEEAVGQ-----EMTIK 207
   | | | | | | : : | : : | : : | : : | : : | : : |
Db 176 YAGIRHSVHLYTPRTPIEEVEVTNLSK-DATGEVVFYSVSVGSAANEADNVLQIQAN 234
QY 208 LGEEBKKI--RTSNRFEVEGEFILIENARFW-----SLEDPLYLPKLELEK-----DEY 253
   | : : : : | : : | : : | : : | : : | : : | : : |
Db 235 LYDKDGLIVANATSDQKLGKLOVNPVKPWWPYLMHSEPGYLQLEIKLATNDELDDVY 294
QY 254 TLDIGIRTISWDEKRLYLNGKPVPLKFGGKHEEPVLGQTFYPLMIKDFNLKWINANS 313
   | : : | : : | : : | : : | : : | : : | : : | : : | : : |
Db 295 RLKVGIRTLTSMNSQOFLINGKPYFRGFRHEDSDIRGKGLDNALMVRDFNLKWIGANA 354
QY 314 FRTHYPYSEEWLIDLADRLGILVIDEAPHVGITRYHYNPETOK-----IA----- 358
   : | | | | | : : | : : | : : | : : | : : | : : |
Db 355 YRTHYPYSEESQFADDEHGIMIDECPVD-TELSYS-DTSKGFLMLLASIFRNFSGE 412
QY 359 -----EDNIRRMIDRHKNHPSVIMWSVANEPEGNHPDAEGFEKALYETANEMDRTRPYVM 413
   : : : : | : | | | : | : | : | : | : | : | : | : | : | : |

```

Db 413 LIGKHKSSLEQLIHRDRNHPVVMWSIANEPRTGSVADSYPFELVANFTRSLDKTRPITA 472

QY 414 VSMNDAPDERTRDVALKYFDIVCNRYGYMYIOGRIEGLQALEKDI EELYA---RHRR 470

Db 473 AIAV---SNTQDKAGRSLDIISFNRYNAMYNSAGRLD---MITQNVIDEAIAWNRKYNK 525

QY 471 PIVFTEFGADAIAGIHYPDPQWSESEYQALVEKTI RL--LLKKDYIIGTHWAFADFK 527

Db 526 PIIMSEYGADILEGLHMOPAYVWSEEFQTEVFSRHKAFDELRRKGWFIIGEFVWNFADEFK 585

QY 528 TPQNVRRPILNHHKGVFTRDRQPKLVAAVLR 558

Db 586 TAQSYTRVGGNKKGVFTRARQPKAAHLRRK 616

RESULT 14  
ADD45479  
ID ADD45479 standard; protein; 648 AA

AC	ADD45479;
XX	
DT	29-JAN-2004 (first entry)
XX	
DE	Rat Protein P06760, SEQ ID NO 10912

KW Rat; pain; neuronal tissue; gene therapy; spinal segmental nerve injury;  
KM chronic constriction injury; CCI; spared nerve injury; SNI; Chung.

OS      Rattus norvegicus.  
XX  
PN      WO2003016475-A2.

PD 27-FEB-2003.

PF 14-AUG-2002; 2002WO-US025765.

PR 14-AUG-2001; 2001US-0312147P.  
PR 01-NOV-2001; 2001US-0346382P.  
PR 26-NOV-2001; 2001US-0333347P.

PA (GEHO ) GEN HOSPITAL CORP.  
PA (FARB ) BAYER AG.

PI Woolf C, D'urso D, Befort K, Costigan M,

DR WPI, 2003-268312/26.  
DR GENBANK; P06760.

**PT** New composition comprising two or more isolated polypeptides, useful for preparing a medicament for treating pain in an animal.

PS Claim 1; Page; 1017pp; English.

CC The invention discloses a composition comprising two or more isolated rat  
CC or human polynucleotides or a polynucleotide which represents a fragment,  
CC derivative or allelic variation of the nucleic acid sequence. Also  
CC claimed are a vector comprising the novel polynucleotide, a host cell  
CC comprising the vector, a method for identifying a nucleotide sequence  
CC which is differentially regulated in an animal subjected to pain and a  
CC kit to perform the method, an array, a method for identifying an agent  
CC that increases or decreases the expression of the polynucleotide sequence  
CC that is differentially expressed in neuronal tissue of a first animal  
CC subjected to pain, a method for identifying a compound which regulates  
CC the expression of a polynucleotide sequence which is differentially  
CC expressed in an animal subjected to pain, a method for identifying a  
CC compound that regulates the activity of one or more of the  
CC polynucleotides, a method for producing a pharmaceutical composition, a  
CC method for identifying a compound or small molecule that regulates the  
CC activity in an animal of one or more of the polypeptides given in the  
CC specification, a method for identifying a compound useful in treating  
CC pain and a pharmaceutical composition comprising the one or more  
CC polypeptides or their antibodies. The polynucleotide or the compound that  
CC modulates its activity is useful for preparing a medicament for treating





QY	204	MTIKLGEEEKKIRTSNRPVEGEFILENARFW----	SLEDP-ylYPLkv-----	ELEKDEY	2533
Db	257	LEVRLLDEDEGKI VARGTGNEGQLKVPRAHLMPYLMHEHPALYLSLEVTMTTPESVSDFY			3166
QY	254	TLDIGIRTISWDEKRLYLNGKPFVLKGFKGHEEFVYLQOGTFYPLMIKDFNLKIMINANS			3133
Db	317	TLPVGIRFVAVTKSKFLLNGKPFYFOGVNKHEDSDLRGRGFDWPLLIKDFNLLRWLGANS			3766
QY	314	FRTSHYPYSEEWLIDLRLGLIVIDEAPHVGIT-----	RYHYNPETOKIAEDNIR		3633
Db	377	FRTSHYPYSEEVLOLCDRYGIVIDECPGVGIVLPOSGFNVSLRHLH----	EVMDDELVR		4311
QY	364	RMIDRHKHPSVIMWSVANEPESNHPDAEGFKALEYETANEMDRTRPVVMWSMMDAPDER			4233
Db	432	---RDKNHPAVVMWSVANEPVSSILKPAGYYFXTLLAHTKALDPTRPVTFVS-----	N		4800
QY	424	TR--DVALKYFDIVCVNRRYYGWYIYQGRIEEGQALIEKDIIEELYARHRKPIFVTEFGAD			4800
Db	481	TRYDADMGAPYVDVICVNSYLSWMYHDYGHLEVIQLOLTSQFENWYKMYQKPIIIOSEYGAD			5400
QY	481	AIAGIHYPDPOMSESEYQALVEKTRILL--	KKDYIIGTHWAFADFKTPQNVRRPIL		5337
Db	541	AVSGLHEDDPRMFSESEYQALLE-NYHLILDEKREYVIGELIWNFADFMNQSPLRVTG			5999
QY	538	NHKGVFTRDRQPKLVAAHVLR-RLW	560		
Db	600	NKKGIPTROPNPKMAAFILRERYW	623		

Search completed: January 24, 2005, 07:50:46  
Job time : 92 secs



QY	361	NIRMRIDRHKHNP	SVIMMSVANE	PESENHPDAEG	FFKALYETANEM	DRTRPVVM	VSMM	DAP	420
Db	361	NIRMRIDRHKHNP	SVIMMSVANE	PESENHPDAEG	FFKALYETANEM	DRTRPVVM	VSMM	DAP	420
QY	421	DERTRDVALK	FPDIVCNR	RYGWIYI	QGRIBEGLOALE	KDIEEL	YARHKRP	FVTE	FGAD
Db	421	DERTRDVALK	FPDIVCNR	RYGWIYI	QGRIBEGLOALE	KDIEEL	YARHKRP	FVTE	FGAD
QY	481	AIAGIHYPDP	QPMFSEEX	QALVEK	TI	RL	LKKDYI	IGTH	WAFADFKTPQNV
Db	481	AIAGIHYPDP	QPMFSEEX	QALVEK	TI	RL	LKKDYI	IGTH	WAFADFKTPQNV
QY	541	GVFTRDRQ	PKLVAH	VLRL	MSEV				
Db	541	GVFTRDRQ	PKLVAH	VLRL	MSEV				

## RESULT 2

```

US-09-270-957-21
; Sequence 21, Application US/09270957
; Patent No. 6641996
; GENERAL INFORMATION:
; APPLICANT: Richard A. Jefferson and Jorge E. Mayer
; TITLE OF INVENTION: MICROBIAL - GLUCURONIDASE GENES, GENE
; TITLE OF INVENTION: PRODUCTS, AND USES THEREOF
; FILE REFERENCE: 190106.405C1
; CURRENT APPLICATION NUMBER: US/09/270,957
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21
; LENGTH: 563
; TYPE: PRT
; ORGANISM: Thermotoga maritima
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(563)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-270-957-21

```

Query Match	99.9%;	Score 2999;	DB 4;	Length 563;
Best Local Similarity	100.0%;	Pred. No. 3e-256;		
Matches 563;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0

QY	1	MVRPQRNKKR	FILILINGVWNL	EVTSKDRPI	IAVPGSMNEQYODL	CTEEGPFTYKTT	PYVPK	60		
Db	1	MVRPQRNKKR	FILILINGVWNL	EVTSKDRPI	IAVPGSMNEQYODL	CTEEGPFTYKTT	PYVPK	60		
QY	61	XL\$QKHIRLY	FAAVNTDCEV	FLNGEKYGENHIEYLP	PEVDVTGKYKSGENEL	RVVVENRL		120		
Db	61	XL\$QKHIRLY	FAAVNTDCEV	FLNGEKYGENHIEYLP	PEVDVTGKYKSGENEL	RVVVENRL		120		
QY	121	KVGGF\$PKVP	BD\$GTHTVG	FFG\$FP	RPNP	DFP	PGGIRPVLIEPTDHARILDI	WDTSES	180	
Db	121	KVGGF\$PKVP	BD\$GTHTVG	FFG\$FP	RPNP	DFP	PGGIRPVLIEPTDHARILDI	WDTSES	180	
QY	181	EPEKKLGKVK	KIEVSEBAVQ	QEMTIKL	G\$EEKIKRTSNR	FVEGEFIL	ENARFWSLED	PY	240	
Db	181	EPEKKLGKVK	KIEVSEBAVQ	QEMTIKL	G\$EEKIKRTSNR	FVEGEFIL	ENARFWSLED	PY	240	
QY	241	LYPLKVEL	EKDEYTL	DIGIRTI	SWDEKRLYL	NGKRVFLK	GF\$KHEBF	PVLGGTFYPLMI	300	
Db	241	LYPLKVEL	EKDEYTL	DIGIRTI	SWDEKRLYL	NGKRVFLK	GF\$KHEBF	PVLGGTFYPLMI	300	
QY	301	KDFNLLK	MINANS	FRTSHYPY	SEEWDL	DADRLG	ILVIDEAPHV	GITRYHYNBETQ	IAED	360
Db	301	KDFNLLK	MINANS	FRTSHYPY	SEEWDL	DADRLG	ILVIDEAPHV	GITRYHYNBETQ	IAED	360
QY	361	NIRRMIDR	HKNHPS	VIWMS	VA	NBPESNHP	DAEGF	KALYETANEMDR	TRPVVMVMSMDAP	420
Db	361	NIRRMIDR	HKNHPS	VIWMS	VA	NBPESNHP	DAEGF	KALYETANEMDR	TRPVVMVMSMDAP	420

[illegible]

### RESULT 3

```

US-09-149-727-2
; Sequence 2, Application US/09149727
; Patent No. 6391547
; GENERAL INFORMATION:
; APPLICANT: Jefferson, Richard A.
; APPLICANT: Kilian, Andrezej
; APPLICANT: Keese, Paul Konrad
; TITLE OF INVENTION: MICROBIAL BETA-GLUCURONIDASE GENES, GENE PRODUCTS AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: 190106.405
; CURRENT APPLICATION NUMBER: US/09/149,727
; CURRENT FILING DATE: 1998-09-08
; EARLIER APPLICATION NUMBER: US 60/058,263
; EARLIER FILING DATE: 1997-09-09
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 602
; TYPE: PRT
; ORGANISM: Bacillus sp.
US-09-149-727-2

```

Query Match	31.0%;	Score 929;	DB 3;	Length 602;
Best Local Similarity	35.6%;	Pred. No. 2.2e-73;		
Matches 221; Conservative	93;	Mismatches 224;	Indels 82;	Gaps 15;

QY	1	MVRPÖRNKKRFLILLINGVMMNLEV-----TSKDRPIAVGSMNE--ÖYÖDLC	44
Dö	1	MLYPINTETRGVPDLNGVMNFEXLDYGKGLBEKMYSKLTDTISMAMPSSSYNDIGUTKEIR	60
QY	45	YEGBFTYKTTFYVPKXLSÖKHTRLFYAAVNNTDCSEVFNLGEKVGENHIEYLPEFVDVTCK	104
Dö	61	NHIGYWMYEREFIVPAYLKDQRVLRFGSATHKAIVYVNGELVEHHKGFLPFEEAINNS	120
QY	105	VKSGENELRVVENRLKVGSPSKVPDSGTHTVFPFGS-----FPPANFDFFPYG	154
Dö	121	JRDGMNRVTVAVDNIL-----DDSTLPYGLJSERHEEGLGVIRKNPNDFFNAYA	170
QY	155	GIRPVLIETDHARILDIWDVTSESEPEREKULGVYKVKI EVSEEA VQEMTIKLG--EEE	212
Dö	171	GUHRPVKIYTTPFTYVEDISVATDENGPR--TGTVTYTVDFO---GKAETVKVSVDDE	223
QY	213	KKIRTSNRFEVGEFILENAREFMSLEDPLYPLKYLEKEDEYTLDI----GIRTISWDEK	267
Dö	224	GKVVASTEGLSGNVEIPNVILMEPLNTLYLQIKVELVMDGLTTIDYEEFPGRVTEVANDG	283
QY	268	RUYLNGKPVELKGFPGKHFEFPVLGÖGTFFPLMKDPNLLKWINANSFRITSHYPYSEEMLD	327
Dö	284	KFLINNKRPFYFKGFGKHEDTPI NGRGFNEASNMVDENILKMWIGANSFRTAHYPYSEEMLR	343
QY	328	LADRLGILVIDEAPHVGITRHYN-----PETÖKA-----EDNIRMIMDR	368
Dö	344	LADREGLVVIDETPAVGV--HLNFMATTGLBGESERSVSTWEKIRTFEHQDVLRRELVSR	400
QY	369	HKNHPSVIMSVANEPESNHDDAEGF KALYETANEMD-RTRPVVMVSMMDAPDERTRDY	427
Dö	401	DKNHPSVVMWSIANEAATEEGAYEYFKPLVELTELKDLPÖKRPTIIVLFVMATPE--TDK	458
QY	428	ALKFYDIVCVNRRYYGWYIYOGRIEEGLÖALEKDIEELYARHR---KPIFVTEGADAIA	483

Db 459 VAEILDIVIALNRNGWYFDGDLBAKVLHR---QEFHAMNKRCPGKIPIMITEYGADTV 515  
QY 484 GIHYDPQPMFSEEQAELEVEKTIRLLKKDYIIGTHWAFADFKTPQNVRRPILNHKGVF 543  
Db 516 GFHDIDPVMFTEEYQVEYYQANHVFDEFENFVGEQAMNFADFATSQGVMRVQGNKKGVF 575  
QY 544 TRDRQPKLVAAHVLRRLMSEV 563  
Db 576 TRDRKPKLAAHVFRERWTNI 595

RESULT 4

US-09-270-957-2  
; Sequence 2, Application US/09270957  
; Patent No. 6641996  
; GENERAL INFORMATION:  
; APPLICANT: Richard A. Jefferson and Jorge E. Mayer  
; TITLE OF INVENTION: MICROBIAL -GLUCURONIDASE GENES, GENE  
; FILE REFERENCE: 190106.405C1  
; CURRENT APPLICATION NUMBER: US/09/270,957  
; CURRENT FILING DATE: 1999-03-17  
; NUMBER OF SEQ ID NOS: 112  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 602  
; TYPE: PRT  
; ORGANISM: Bacillus sp.  
US-09-270-957-2

Query Match 31.0%; Score 929; DB 4; Length 602;  
Best Local Similarity 35.6%; Pred. No. 2.2e-73;  
Matches 221; Conservative 93; Mismatches 224; Indels 82; Gaps 15;

QY 1 MVRPQRNKKRFILINGVWNLEV-----TSKDRPIAVPGSWNE--QYQDLC 44  
Db 1 MLYPINTETRGVFDLNGVWNFKLDYKGLEEKWYESKLTDTISMAVPSSYNDIGVTKAIR 60  
QY 45 YEEGPFTYKTFYVPKXLSQKHRLYFAAVNTDCEVFLNGEKGVENHIEYLPEYDVTGK 104  
Db 61 NHIGVWYEREFVPAVLKQRIYLRFGSATHKAIYVNGELVVEHKGGLPFPEAEINNS 120  
QY 105 VKSGENELRVVVENRLKVGFPSPKVPDSGTHYGFSGS-----PPRANDFPPYG 154  
Db 121 LRDGMNRVTVAVDNIL-----DSTLPGLYSERHEEGLGKVI RNKPNDFENYA 170  
QY 155 GIIRPVLIEFTDHARILDIWDTSESEPEKKLGKVKKIEVSEAVQEMTIKLG--EEB 212  
Db 171 GHRPVKIYTTPTTYVEDISVTDENGPR--TGTVYTVDFQ----GKAETVKSVDDE 223  
QY 213 KKIIRTSNRFVEGEFILENARFWSLEDPYLPKVELEKDEYTLDI-----GIRTSWDEK 267  
Db 224 GKVVASTEGLSGWEIIPNVILMEPLNTLYLQIKVELVNDGLTIDYEEFPGVRTVEVNDG 283  
QY 268 RLYLNGKPVFLKGFKGHEEPVLGQGTFFPLMIKDNLLKWINANSFRTSHYPYSEEWLD 327  
Db 284 KFLINNKPFYFGFGKHEDTPINGRGENASNVMDFNILKWIGANSFRTAHYPSSELMR 343  
QY 328 LADRLGILVIDEAPHVGITRHYN-----PETOKIA-----EDNIRRMIDR 368  
Db 344 LADREGLVVIDETPAVG---HLNFMATTGLGEGSERVSTWEKIRTFEHHQDVLRVSR 400  
QY 369 HKNHPSVIMSVANEPESNHPDAEGFKALYETANEMD-RTRPVVMSMDADERTRDV 427  
Db 401 DKNHPSVVMWSIANEAATEEGAYEFKPLVELTKELDPQKRPTIVLFVMAPE--TDK 458  
QY 428 ALKYFDIVCVNRYGYWYIYQGRIEGLQALEKDIIEELYARHR---KPIFVTEGADAIA 483  
Db 459 VAEILDIVIALNRNGWYFDGDLBAKVLHR---QEFHAMNKRCPGKIPIMITEYGADTV 515  
QY 484 GIHYDPQPMFSEEQAELEVEKTIRLLKKDYIIGTHWAFADFKTPQNVRRPILNHKGVF 543  
Db 576 TRDRKPKLAAHVFRERWTNI 595

Db 516 GFHDIDPVMFTEEYQVEYYQANHVFDEFENFVGEQAMNFADFATSQGVMRVQGNKKGVF 575  
QY 544 TRDRQPKLVAAHVLRRLMSEV 563  
Db 576 TRDRKPKLAAHVFRERWTNI 595

RESULT 5

US-09-270-957-8  
; Sequence 8, Application US/09270957  
; Patent No. 6641996  
; GENERAL INFORMATION:  
; APPLICANT: Richard A. Jefferson and Jorge E. Mayer  
; TITLE OF INVENTION: MICROBIAL -GLUCURONIDASE GENES, GENE  
; FILE REFERENCE: 190106.405C1  
; CURRENT APPLICATION NUMBER: US/09/270,957  
; CURRENT FILING DATE: 1999-03-17  
; NUMBER OF SEQ ID NOS: 112  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 8  
; LENGTH: 602  
; TYPE: PRT  
; ORGANISM: Bacillus sp.  
US-09-270-957-8

Query Match 31.0%; Score 929; DB 4; Length 602;  
Best Local Similarity 35.6%; Pred. No. 2.2e-73;  
Matches 221; Conservative 93; Mismatches 224; Indels 82; Gaps 15;

QY 1 MVRPQRNKKRFILINGVWNLEV-----TSKDRPIAVPGSWNE--QYQDLC 44  
Db 1 MLYPINTETRGVFDLNGVWNFKLDYKGLEEKWYESKLTDTISMAVPSSYNDIGVTKAIR 60  
QY 45 YEEGPFTYKTFYVPKXLSQKHRLYFAAVNTDCEVFLNGEKGVENHIEYLPEYDVTGK 104  
Db 61 NHIGVWYEREFVPAVLKQRIYLRFGSATHKAIYVNGELVVEHKGGLPFPEAEINNS 120  
QY 105 VKSGENELRVVVENRLKVGFPSPKVPDSGTHYGFSGS-----PPRANDFPPYG 154  
Db 121 LRDGMNRVTVAVDNIL-----DSTLPGLYSERHEEGLGKVI RNKPNDFENYA 170  
QY 155 GIIRPVLIEFTDHARILDIWDTSESEPEKKLGKVKKIEVSEAVQEMTIKLG--EEB 212  
Db 171 GHRPVKIYTTPTTYVEDISVTDENGPR--TGTVYTVDFQ----GKAETVKSVDDE 223  
QY 213 KKIIRTSNRFVEGEFILENARFWSLEDPYLPKVELEKDEYTLDI-----GIRTSWDEK 267  
Db 224 GKVVASTEGLSGWEIIPNVILMEPLNTLYLQIKVELVNDGLTIDYEEFPGVRTVEVNDG 283  
QY 268 RLYLNGKPVFLKGFKGHEEPVLGQGTFFPLMIKDNLLKWINANSFRTSHYPYSEEWLD 327  
Db 284 KFLINNKPFYFGFGKHEDTPINGRGENASNVMDFNILKWIGANSFRTAHYPSSELMR 343  
QY 328 LADRLGILVIDEAPHVGITRHYN-----PETOKIA-----EDNIRRMIDR 368  
Db 344 LADREGLVVIDETPAVG---HLNFMATTGLGEGSERVSTWEKIRTFEHHQDVLRVSR 400  
QY 369 HKNHPSVIMSVANEPESNHPDAEGFKALYETANEMD-RTRPVVMSMDADERTRDV 427  
Db 401 DKNHPSVVMWSIANEAATEEGAYEFKPLVELTKELDPQKRPTIVLFVMAPE--TDK 458  
QY 428 ALKYFDIVCVNRYGYWYIYQGRIEGLQALEKDIIEELYARHR---KPIFVTEGADAIA 483  
Db 459 VAEILDIVIALNRNGWYFDGDLBAKVLHR---QEFHAMNKRCPGKIPIMITEYGADTV 515  
QY 484 GIHYDPQPMFSEEQAELEVEKTIRLLKKDYIIGTHWAFADFKTPQNVRRPILNHKGVF 543  
Db 576 TRDRKPKLAAHVFRERWTNI 595

```
RESULT 6
US-09-270-957-15
; Sequence 15, Application US/09270957
; Patent No. 6641996
; GENERAL INFORMATION:
; APPLICANT: Richard A. Jefferson and Jorge E. Mayer
; TITLE OF INVENTION: MICROBIAL - GLUCURONIDASE GENES, GENE
; TITLE OF INVENTION: PRODUCTS, AND USES THEREOF
; FILE REFERENCE: 190106.405C1
; CURRENT APPLICATION NUMBER: US/09/270,957
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15
; LENGTH: 602
; TYPE: PRT
; ORGANISM: Bacillus sp.
; US-09-270-957-15
```

```
Query Match          31.0%; Score 929; DB 4; Length 602;
Best Local Similarity 35.6%; Pred. No. 2.2e-73;
Matches 221; Conservative 93; Mismatches 224; Indels 82; Gaps 15;

QY 1 MVRPQRNKKRFFILNGVWNLV-----TSKDRPIAVPGSWNE--QYQDLC 44
D 1 MLYPINTETRGVFDLNGVWNLFDYKGLBEKWYESKLTDTISMAVPSYNDIGVTKR 60
QY 45 YEEGPFTYKTFYVVPKXLSQKHIRLYFAAVNTDCEVFLNGEKGVENHIEYLPFEVDVTGK 104
D 61 NHIGYVWEREFVPAVYLKDQRIVLRFGSATHKALIVYNGELVVEHKGGLPFEAEINNS 120
QY 105 VKSGENELRVVENRLKVGFPSPKVPDSGTHTVGFFGS-----FPPANPDFPFG 154
D 121 LRDGMNRVTVAVDNL-----DSTLPVGLYSEHHEGLGKVI RNKPNDFFNVA 170
QY 155 GIIRPVLIETDHA RILDIWDTSESEPEKKLGKVKIEVSEEA VGOEMTIKLG--EEB 212
D 171 GLHRPVKIYTPFTYVEDISVTDENGPR--TGTVTYTVDFQ----GKAETVKVSVDEE 223
QY 213 KKI RTSNRFEVEGEFLENNARFWSLEDPLYPLKVELEKDEYTLDI-----GIRTSWDEK 267
D 224 GKVA STEGLSGNVEIPNVILMEPLNTYLYQIKVELVNDGLTIDVYEEBPGVRTVEVNDG 283
QY 268 RLYLNGKPVFLKGFQKHEEPVVLGQGTFFPLMIKDFNLKWINANSFRTSHYPYSEEWLD 327
D 284 KFLINNKPFYFKGFGKHEDTPINGRGFNEASNVDFNLKWINANSFRTAHYPYSEELMR 343
QY 328 LADRLGLIVIDEAPHVIGITRYHYN-----PETOKIA-----EDNIRRMIDR 368
D 344 LADREGLVVIDETPAVG---HLNFMATTLGEGSERVSTWEKIRTFEHHQDVLRELVS 400
QY 369 HKNHPSVIMSVANEPESNHPDAEGFFKALYETANEMD--RTRPVVMVSMMDAPDERTDV 427
D 401 DKNHPSVVMWSIANEAATEEGAYEYFKPLVELTKELDPQKRPTIVLFVMAATPE--TDK 458
QY 428 ALKYFDIVCVNRYGYMYIYQGRIEBGLQALEKDIIEELYARHR---KPIFVTEFGADAIA 483
D 459 VAE LIDVIALNRNGWYFPGDLEAAKVHLR---QEFHAMNKRCPGPKIMITEYGADTV 515
QY 484 GIHYDPQWFSSEYQAEIVKTI RLLKKDYIIGTHVMAFADFKTPQNVRRPILNKGVF 543
D 516 GFHDIDPVMFTEEYQVEYYQANHVFDEFENFVGEQAMNFPADFATISQGVMRVQGNKGV 575
QY 544 TRDRQPKLVAVHLRLMSEV 563
D 576 TRDRKPKLAHVFRERWTNI 595
```

```
RESULT 7
US-09-149-727-4
; Sequence 4, Application US/09149727
```

```
; Patent No. 6391547
; GENERAL INFORMATION:
; APPLICANT: Jefferson, Richard A.
; APPLICANT: Kilian, Andrzej
; APPLICANT: Keese, Paul Konrad
; TITLE OF INVENTION: MICROBIAL BETA-GLUCURONIDASE GENES, GENE PRODUCTS AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: 190106.405
; CURRENT APPLICATION NUMBER: US/09/149,727
; CURRENT FILING DATE: 1998-09-08
; EARLIER APPLICATION NUMBER: US 60/058,263
; EARLIER FILING DATE: 1997-09-09
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 618
; TYPE: PRT
; ORGANISM: Bacillus sp.
; US-09-149-727-4
```

```
Query Match          31.0%; Score 929; DB 3; Length 618;
Best Local Similarity 35.6%; Pred. No. 2.3e-73;
Matches 221; Conservative 93; Mismatches 224; Indels 82; Gaps 15;

QY 1 MVRPQRNKKRFFILNGVWNLV-----TSKDRPIAVPGSWNE--QYQDLC 44
D 17 MLYPINTETRGVFDLNGVWNLFDYKGLBEKWYESKLTDTISMAVPSYNDIGVTKR 76
QY 45 YEEGPFTYKTFYVVPKXLSQKHIRLYFAAVNTDCEVFLNGEKGVENHIEYLPFEVDVTGK 104
D 77 NHIGYVWEREFVPAVYLKDQRIVLRFGSATHKALIVYNGELVVEHKGGLPFEAEINNS 136
QY 105 VKSGENELRVVENRLKVGFPSPKVPDSGTHTVGFFGS-----FPPANPDFPFG 154
D 137 LRDGMNRVTVAVDNL-----DSTLPVGLYSEHHEGLGKVI RNKPNDFFNVA 186
QY 155 GIIRPVLIETDHA RILDIWDTSESEPEKKLGKVKIEVSEEA VGOEMTIKLG--EEB 212
D 187 GLHRPVKIYTPFTYVEDISVTDENGPR--TGTVTYTVDFQ----GKAETVKVSVDEE 239
QY 213 KKI RTSNRFEVEGEFLENNARFWSLEDPLYPLKVELEKDEYTLDI-----GIRTSWDEK 267
D 240 GKVA STEGLSGNVEIPNVILMEPLNTYLYQIKVELVNDGLTIDVYEEBPGVRTVEVNDG 299
QY 268 RLYLNGKPVFLKGFQKHEEPVVLGQGTFFPLMIKDFNLKWINANSFRTSHYPYSEEWLD 327
D 300 KFLINNKPFYFKGFGKHEDTPINGRGFNEASNVDFNLKWINANSFRTAHYPYSEELMR 359
QY 328 LADRLGLIVIDEAPHVIGITRYHYN-----PETOKIA-----EDNIRRMIDR 368
D 360 LADREGLVVIDETPAVG---HLNFMATTLGEGSERVSTWEKIRTFEHHQDVLRELVS 416
QY 369 HKNHPSVIMSVANEPESNHPDAEGFFKALYETANEMD--RTRPVVMVSMMDAPDERTDV 427
D 417 DKNHPSVVMWSIANEAATEEGAYEYFKPLVELTKELDPQKRPTIVLFVMAATPE--TDK 474
QY 428 ALKYFDIVCVNRYGYMYIYQGRIEBGLQALEKDIIEELYARHR---KPIFVTEFGADAIA 483
D 475 VAE LIDVIALNRNGWYFPGDLEAAKVHLR---QEFHAMNKRCPGPKIMITEYGADTV 531
QY 484 GIHYDPQWFSSEYQAEIVKTI RLLKKDYIIGTHVMAFADFKTPQNVRRPILNKGVF 543
D 532 GFHDIDPVMFTEEYQVEYYQANHVFDEFENFVGEQAMNFPADFATISQGVMRVQGNKGV 591
QY 544 TRDRQPKLVAVHLRLMSEV 563
D 592 TRDRKPKLAHVFRERWTNI 611
```

```
RESULT 8
US-09-149-727-8
; Sequence 8, Application US/09149727
; Patent No. 6391547
```





```

; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 28
; LENGTH: 615
; TYPE: PRT
; ORGANISM: Bacillus sp.
US-09-270-957-28

```

Query Match	30.8%;	Score 925;	DB 4;	Length 615;
Best Local Similarity	35.7%;	Pred. No. 5.2e-73;		
Matches 220;	Conservative 92;	Mismatches 223;	Indels 82;	Gaps 15;

2y	4	POBNKKRFILILNGVWNLV-----TSXDRPIAVPGSNNE--QYQDLCTYEE	47
		: : :         : : :   :     : : :   : :	
Db	9	PINTETRGVFDLNGVWNLFKLDYGKGLBEKMYESKLTDTISMAVPPSSINDIGVTKEIRNHI	68
Qy	48	GPFYTYKTFYVPKXLSQKHIRLYFAAVNTDCEVFLNGEKVGENHIEYLPFEVDVTGKYKS	107
		: : :       : : :   : : :   :       : : :   : :	
Db	69	GYWYEREFVTPAYLKQDRIVLRFSGATHKAIIVYNGELVVEHKGGFLPFEAEINNSLRD	128
Qy	108	GENELRVVENRLKYGGFSPSKVPDSCGTHTVGFFGS-----FPANDFFPYGGII	157
		: : :   :   : : :   : :   : : :   :       : :	
Db	129	GNNRVTVAVDNIL-----DDSLPLVGLYSERHEGLGKVI RNKNPFDFFNYAGLH	178
Qy	158	RPVLIEFTDHARILDIWDTSESEPEKKLGKVKVKEVSEAVGQEMTIKLG--EEEKKI	215
		: : :   :   : : :   : : :   : : :   : : :	
Db	179	RPVKIYTTPTTYVEDISVTTDNGP--TGTVTYTVBQ---GKAETVKVSVVDEEGKV	231
Qy	216	RTSNRFVEGEFILENARFWSLEDPYLPKVELKEDXYLDI----GIRTISWDEKRLY	270
		: : :   : : :   : :   :   :   : :   : :   : :	
Db	232	VASTEGLSGNVEIPNVILMEPLNTYLYQIKVELVNDGLTIDVYEEPFQVRTVEVNDGKEL	291
Qy	271	LNKGPVFLKGFKGHEEFVPLGQGTFPYPLMIKD FNILKWINANSFRTSHYPYSEEWLIDLAD	330
		: : :         : : :   : : :   :               : :	
Db	292	INNRPYFKGFGKHEDTPINGRGFNEASNVMDFNILKIGANSFRTAHYPYSEELMRILAD	351
Qy	331	RLGLIVIDEAPHVGTITRYHYN-----PETOKIA-----EDNIRRMIDRHKN	371
		: :         : :   : :   : :   : :   : :   : :   : :	
Db	352	REGIIVIDETPAVGV---HLNFMAITGEGESERSVSTWEKIRTFEHHDVLRRELSRDXN	408
Qy	372	HPSVIMSVANEPESNHPDAEGFFKALYETANEMD-RTRPVMVMSMDAPDERTRDVALK	430
		: :       : : :   : :   : :   : :   : :   : :   : :	
Db	409	HPSVVMWSIANEAATEEGAYEYFKPLVELTKELDPQKRPVTIVLFVMAYPE--TDKYAE	466
Qy	431	YFDIVCNRRYYGYIYQGRIEBGLQALEKDI EELYARHR---KPIFVTEFGADAIAGIH	486
		: :         : :   : :   : : :   : : :   : :   : :	
Db	467	LIDIALNRRYNGWYFDGGDLLEAKVHLR---QEFHANMRCPGKPRIMITEYGADTVAGFH	523
Qy	487	YDPPOMFSEEYQAEIVEXTIRLLKKDYIIGTHVWAPADFKTPQNVRRPILNHKGVFTRD	546
		: :         : : :   : : :   :               : :	
Db	524	DIDPMVFTEEYQVEYYQANHVFDFEENFVGEQAMNFPADFATISQGVMRVQGNKKGVFTRD	583
Qy	547	ROPKLVAHVLRRLMSEV	563
		:             : : :   : :   : :   : :   : :	
Db	584	RKPKLAHVFRERWTNI	600

```

RESULT 11
; US-09-862-660-2
; Sequence 2, Application US/09862660
; Patent No. 6664097
; GENERAL INFORMATION:
; APPLICANT: Russell, William
; APPLICANT: Klaenhammer, Todd
; TITLE OF INVENTION: LACTOBACILLUS BETA-GLUCURONIDASE AND DNA ENCODING THE SAME
; FILE REFERENCE: 5051.514
; CURRENT APPLICATION NUMBER: US/09/862,660
; CURRENT FILING DATE: 2001-05-21
; PRIOR APPLICATION NUMBER: 60/206,372
; PRIOR FILING DATE: 2000-05-23
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.0

```

```
; SEQ ID NO 2
;
; LENGTH: 598
; TYPE: PRT
; ORGANISM: Lactobacillus gasseri
;
US-09-862-660-2
```

Query Match	30.6%;	Score 919.5;	DB 4;	Length 598;
Best Local Similarity	34.1%;	Pred. No. 1.5e-72;		
Matches 207; Conservative	110;	Mismatches 227;	Indels 63;	Gaps 14;

[illegible]

```

RESULT 12
US-09-715-858-4
; Sequence 4, Application US/09715858
; Patent No. 6582692
; GENERAL INFORMATION:
; APPLICANT: Podsakoff, Gregory
; APPLICANT: Watson, Gordon
; APPLICANT: Couto, Linda B.
; APPLICANT: Yang, Bin
; TITLE OF INVENTION: RECOMBINANT ADENO-ASSOCIATED VIRUS VIRIONS FOR THE
; TITLE OF INVENTION: TREATMENT OF LYSOSOMAL DISORDERS
; FILE REFERENCE: 0800-0021
; CURRENT APPLICATION NUMBER: US/09/715, 858
; CURRENT FILING DATE: 2000-11-14
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 648
;

```

TYPE: PRT  
ORGANISM: Mus musculus  
US-09-715-858-4

Query Match 30.0%; Score 899.5; DB 4; Length 648;  
Best Local Similarity 35.9%; Pred. No. 1e-70;  
Matches 224; Conservative 101; Mismatches 208; Indels 91; Gaps 21;

```
QY 1 MVRPQRNKKRPILILNGVWNLV-TSKDR-----PIAVGSMNEQ 39
DB 27 MLFPKSPSRELKALDGLMHRADLSNNRLOGFEQQWYRQPLRESGPVLDMPVPSSFNDI 86
QY 40 YQDLCYEE--GPFTYKTFYVPKLSQ--KHIRLYFAAVNTDCEVFLNGEKVGENHIEY 94
DB 87 TQEALRDFIGWVYEREAIRPRWTQDTDMRVLRINSAHYAAVWVWNGIHVEHEGCH 146
QY 95 LPFEVDVTGKYKSG--ENELRVVENRLKYGGFPSPKVP-----DSGTHVGFSGSF 143
DB 147 LPFEADISKLVQSGPLTTCRTITAINNTLT---PHTLPPGTIVYKTDTSWYPKGYF--V 200
QY 144 PPANFDFFPYGGIIRPVLIETDHAIRLDIWDTSESEPEKKLGKVK--VKIEVSEAVG 201
DB 201 QDTSFDFFNAGLHRSVLYTPTTYIDITVITN--VEQDIGLVTYWISVQSEHF-- 255
QY 202 QEMTIKLGEEBKIRTSNRFVEGEFILENARFW---SLEDP-VLYPLKV-----ELEKD 251
DB 256 -QLEVQLDEGGKVVAHGTGNGOQLQVPSANLWMPYLMHEHPAYMSLEVKYTTTESVTD 314
QY 252 EYTLDIGIRISWDEKRLYLNGKPVFLKFGKHEEPVLGQGTFFYPLMIKDFNLKWINA 311
DB 315 YTTLPIGIRIVAVTKSKFLINGRPFYFGVNGHEDSDIRGKFDWPLVTKDFNLRWLGA 374
QY 312 NSFRTSHYPYSEEWLADRLGILVIDEAPHVGIT-----RYHYNPETQKIAEDN 361
DB 375 NSFRTSHYPYSEEVQLCDRYGIVVIDECPGVGIVLPQSGNESLRHHL-----EVMEEL 429
QY 362 IRRMIDRKHNPVSIMSVANBESNHPDAEGFFKALYETANEMDRTRPVVMVS--MMDA 419
DB 430 VR---RDKNHPAVVMWSVANBESSALKRPAAYEFKTLITHTKALDLTRPVTEFVSNKYDA 485
QY 420 PDERTDVALKYFDIVCVNRYGYIYQGRIEGLQALEKDIIEELYARHKRPIFTEFGA 479
DB 486 -----DLGAPYVDVICVNSYFSWYHDYGHLEVIQPOLNSQFENWYKTHQKPIIQSEYGA 539
QY 480 DALAGIHYPDPQMFSEEQALVEKTIRLL--LKDDYIIGTHVAFADFKTPQNVRRPIL 537
DB 540 DAIPGIHEDPRMFSEEQKAVLENHYSVLDQKRKEVYVGGELIWNFADFMTNQSPLRVIG 599
QY 538 NHKGVFTDRQPKLVAVHLR-RLW 560
DB 600 NKGIFTRQRPKTSAFILRERYW 623
```

RESULT 13  
US-09-118-276-12

Sequence 12, Application US/09118276

Patent No. 6693185

GENERAL INFORMATION:

APPLICANT: BABYCHUK, ELENA;

APPLICANT: KUSHNIR, SERGEI;

APPLICANT: DE BLOCK, MARC;

APPLICANT: INZE, DIRK

TITLE OF INVENTION: METHODS AND MEANS TO MODULATE PROGRAMMED

TITLE OF INVENTION: CELL DEATH IN EUKARYOTIC CELLS

NUMBER OF SEQUENCES: 21

CORRESPONDENCE ADDRESS:

ADDRESSEE: SIBBEY, FRIEDMAN, LEEDOM, & FERGUSON

STREET: 8180 GREENSBORO DRIVE, SUITE 800

CITY: MCLEAN,

STATE: VIRGINIA

COUNTRY: USA

ZIP: 22102

COMPUTER READABLE FORM:

```
MEDIUM TYPE: 3-1/2" DISKETTE
COMPUTER: IBM-COMPATIBLE
OPERATING SYSTEM: MS-DOS
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/118,276
FILING DATE: 17-JUL-1998
PRIOR APPLICATION DATA: NONE
ATTORNEY/AGENT INFORMATION:
NAME: SCHULMAN, ROBERT M.; SALEN, KENNETH H.
REGISTRATION NUMBER: 31,196; 43,077
REFERENCE/DOCKET NUMBER: 6201-0003
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 790-9110
TELEFAX: (703) 883-0370
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 1010 RESIDUES
TYPE: AMINO ACID
STRANDEDNESS: SINGLE
TOPOLOGY: LINEAR
US-09-118-276-12
```

Query Match 29.4%; Score 882; DB 4; Length 1010;  
Best Local Similarity 34.0%; Pred. No. 6.9e-69;  
Matches 206; Conservative 109; Mismatches 229; Indels 62; Gaps 11;

```
QY 1 MVRPQRNKKRPILILNGVWNLV-TSKDRPIAVPGSMNEQYQDLCY 45
DB 408 MVRPVEPTFREIKKLDGLMAFSLDRENGCIGDQRMWESALQESRAIAVPGSFNDQFADADI 467
QY 46 EE--GPFTYKTFYVPKLSQKHIRLYFAAVNTDCEVFLNGEKVGENHIEYLPFEVDVTG 103
DB 468 RNYAGNVWYQREVFIPKGMACQRIVLRFDVATHYKQVWVNNQEVMEHQGGYTPFEADVTP 527
QY 104 KYKSGEN-ELRVVENRLKYGGFPSPKV---DSGTHVGFSGFPPANFDFFPYGGIIRP 159
DB 528 YVIAGKSVRITVCVNNELNQITPPGMVITDENGKKQSYF-----HDFPNYAGIHS 580
QY 160 VLIBTDHARILDIWDTSESEPEKKLGKVKYKIEVSEAVQEMTIKLGEEBKIRTSN 219
DB 581 VMLYTPNTWDDITVTVHVAQ--DCNHASVDWQV--VANGDVSVELRDADQOVAVTG 634
QY 220 RFVEGEFILENARFWSLEDPYLPYPL---KVELEKDEYTLDIGIRISWDEKRLYLNGKP 275
DB 635 QGTSGLQVNNPMLMQPGEGLYELCVTAKSQTECDIYPLRIGIRSVAVKGGQFLINHKP 694
QY 276 VFLKFGKHEEPVLGQGTFFYPLMIKDFNLKWINANSFRTSHYPYSEEWLADRLGIL 335
DB 695 FYFTGFRHEDADLRGKGFNVLMVHDHALMDWIGANSYRTSHYPYAEEMLDWADEHGIV 754
QY 336 VIDEAPHVGI-----TRYHYNPETQKIAEDNIRRMIDRKHNPVSIM 377
DB 755 VIDETAAVGNLISLIGIFEAGNKKPKELYSEAVNGETQAHLOAIKELIARDKNHPŠVVM 814
QY 378 MSVANEPESNHPDAEGFFKALYETANEMDRTRPVVMVSMW--DAPDERTDVALKYFDIV 435
DB 815 WSIANEPDTRPQAREYFAPLAETRKLDPTRPITCVNVWMCDAHTDTISDL---FDVL 870
QY 436 CVNRYYGWYIYQGRIEEGLQALEKDIIEELYARHKRPIFTEFGADAIAGIHYPDPQMFSE 495
DB 871 CLNRYYGWYQSGDLTAEKYLEKELLAWQEKLHQPIITFYGVDTLAIGLSMYTDMWSE 930
QY 496 EYQALVEKTIRLLKKDYIIGTHVAFADFKTPQNVRRPILNHKGVFTDRQPKLVAVH 555
DB 931 EYQCAWLDMYHRVFDRAVAVGEQVWVNFADPATSGILRVGNKKGIFTRDRKPKSAAPL 990
QY 556 LRLMS 561
DB 991 LQKRW 996
```

RESULT 14

```
US-09-149-727-6
; Sequence 6, Application US/09149727
; Patent No. 6391547
; GENERAL INFORMATION:
; APPLICANT: Jefferson, Richard A.
; APPLICANT: Kilian, Andrzej
; APPLICANT: Keese, Paul Konrad
; TITLE OF INVENTION: MICROBIAL BETA-GLUCURONIDASE GENES, GENE PRODUCTS AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: 190106.405
; CURRENT APPLICATION NUMBER: US/09/149,727
; CURRENT FILING DATE: 1998-09-08
; EARLIER APPLICATION NUMBER: US 60/058,263
; EARLIER FILING DATE: 1997-09-09
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 603
; TYPE: PRT
; ORGANISM: Escherichia coli
; US-09-149-727-6

Query Match      29.3%; Score 879; DB 3; Length 603;
Best Local Similarity 33.8%; Pred. No. 5.8e-69;
Matches 205; Conservative 110; Mismatches 229; Indels 62; Gaps 11;

QY      1 MVRPQRNKKRFFILLNGVWNLV-----TSKDRPIAVPGSWNEQYDLCY 45
DB      1 MLRPVETPTREIKKLDGLMAFSLDRENGCIDQRWESALQESRAIAVPGSFNDQFADADI 60
QY      46 EE--GPFTYKTTFFVVPKXLSQKHIRLYFAAVNTDCEVFLNGEKVGENHIEYLPFEVDVTG 103
DB      61 RNYAGNVWYQREVFIPKMGAGORIVLRFDAVTHYGKVVWVNNQEVMEHOGGYTPFEADVTP 120
QY      104 KVKSGEN-ELRVVVENRLKVGSPSKV---PDSGTHVGFSGFPANFDFPYGGIIRP 159
DB      121 YVIAGKSVRITVCVNNELNWQTIIPGMVITDENGKKQSYF-----HDFNYAGIHR 173
QY      160 VLIEFTDHARILDIWDTSESEPEKKLGKVKYKIEVSEEAVGQEMTIKLGEEKKIRTSN 219
DB      174 VMLYTPNTWDDITVTHVAQ---DCNHSVDMQV---VANGDVSVELRDADQOVVATG 227
QY      220 RFVEGEFILENARFWSLEDPYLPYPL---KVELEKDEYTLDIGRTISWDEKRLYLNGKP 275
DB      228 QGTSGLQVNVPHLMQPGEGYLYELCVTAKSQTECDIYPLRVGIRSAVAVKGEQFLINHKP 287
QY      276 VFLKGFGEKHEEFVPLGQGTFFPLMIKDFNLKWINANSFRTSHYPYSEEWLDLADRLGIL 335
DB      288 FYFTGFRHEDADLRGKGFNDVLMVHDHLMWIGANSYRTSHYPYAEEMLDWADHEGIV 347
QY      336 VIDEAPHVGI-----TRYHYNPETOKIAEDNIRRMIDRHKNHPSVIM 377
DB      348 VIDETAAGFNLSLGIGFEAGNKKPKELYSEEA VNGETQQAHLQAIKELIARDKNHPSVIM 407
QY      378 WSVANEPESNHPDAEGFFKALYETANEMDRTRPVVMVSM--DAPDERTRDVALKYFDIV 435
DB      408 WSIANEDTRPQAGAREYFAPLAETARKLDPTRPITCVNVVMFCDAHTDTISDL---FDVL 463
QY      436 CVNRYYGWYIYQGRIEEGLQALEKDIIEELVARRHKPIFVTEFGADAIAGIHYDPQMFSE 495
DB      464 CLNRYYGWYVQSGDLETAEKVLEKELLAWQEKLHQPIIITEYGVDTLAGLHSMYTTDMWSE 523
QY      496 EYQAEVLEKTIIRLLKKDYIIGTHVWAFADFKTPQNVRRPILNHKGVFTRDRQPKLVAVH 555
DB      524 EYQCAWLDMYHRVFDRAVSAVVEQVWVNFADFATSGILRVGNGKKGIFTRDRKPKSA AFL 583
QY      556 LRLRWS 561
DB      584 LQKRWT 589

RESULT 15
US-09-270-957-17
```

```
; Sequence 17, Application US/09270957
; Patent No. 6641996
; GENERAL INFORMATION:
; APPLICANT: Richard A. Jefferson and Jorge E. Mayer
; TITLE OF INVENTION: MICROBIAL -GLUCURONIDASE GENES, GENE
; TITLE OF INVENTION: PRODUCTS, AND USES THEREOF
; FILE REFERENCE: 190106.405C1
; CURRENT APPLICATION NUMBER: US/09/270,957
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: FastSeq for windows Version 4.0
; SEQ ID NO 17
; LENGTH: 603
; TYPE: PRT
; ORGANISM: Escherichia coli
; US-09-270-957-17

Query Match      29.3%; Score 879; DB 4; Length 603;
Best Local Similarity 33.8%; Pred. No. 5.8e-69;
Matches 205; Conservative 110; Mismatches 229; Indels 62; Gaps 11;

QY      1 MVRPQRNKKRFFILLNGVWNLV-----TSKDRPIAVPGSWNEQYDLCY 45
DB      1 MLRPVETPTREIKKLDGLMAFSLDRENGCIDQRWESALQESRAIAVPGSFNDQFADADI 60
QY      46 EE--GPFTYKTTFFVVPKXLSQKHIRLYFAAVNTDCEVFLNGEKVGENHIEYLPFEVDVTG 103
DB      61 RNYAGNVWYQREVFIPKMGAGORIVLRFDAVTHYGKVVWVNNQEVMEHOGGYTPFEADVTP 120
QY      104 KVKSGEN-ELRVVVENRLKVGSPSKV---PDSGTHVGFSGFPANFDFPYGGIIRP 159
DB      121 YVIAGKSVRITVCVNNELNWQTIIPGMVITDENGKKQSYF-----HDFNYAGIHR 173
QY      160 VLIEFTDHARILDIWDTSESEPEKKLGKVKYKIEVSEEAVGQEMTIKLGEEKKIRTSN 219
DB      174 VMLYTPNTWDDITVTHVAQ---DCNHSVDMQV---VANGDVSVELRDADQOVVATG 227
QY      220 RFVEGEFILENARFWSLEDPYLPYPL---KVELEKDEYTLDIGRTISWDEKRLYLNGKP 275
DB      228 QGTSGLQVNVPHLMQPGEGYLYELCVTAKSQTECDIYPLRVGIRSAVAVKGEQFLINHKP 287
QY      276 VFLKGFGEKHEEFVPLGQGTFFPLMIKDFNLKWINANSFRTSHYPYSEEWLDLADRLGIL 335
DB      288 FYFTGFRHEDADLRGKGFNDVLMVHDHLMWIGANSYRTSHYPYAEEMLDWADHEGIV 347
QY      336 VIDEAPHVGI-----TRYHYNPETOKIAEDNIRRMIDRHKNHPSVIM 377
DB      348 VIDETAAGFNLSLGIGFEAGNKKPKELYSEEA VNGETQQAHLQAIKELIARDKNHPSVIM 407
QY      378 WSVANEPESNHPDAEGFFKALYETANEMDRTRPVVMVSM--DAPDERTRDVALKYFDIV 435
DB      408 WSIANEDTRPQAGAREYFAPLAETARKLDPTRPITCVNVVMFCDAHTDTISDL---FDVL 463
QY      436 CVNRYYGWYIYQGRIEEGLQALEKDIIEELVARRHKPIFVTEFGADAIAGIHYDPQMFSE 495
DB      464 CLNRYYGWYVQSGDLETAEKVLEKELLAWQEKLHQPIIITEYGVDTLAGLHSMYTTDMWSE 523
QY      496 EYQAEVLEKTIIRLLKKDYIIGTHVWAFADFKTPQNVRRPILNHKGVFTRDRQPKLVAVH 555
DB      524 EYQCAWLDMYHRVFDRAVSAVVEQVWVNFADFATSGILRVGNGKKGIFTRDRKPKSA AFL 583
QY      556 LRLRWS 561
DB      584 LQKRWT 589
```

Search completed: January 24, 2005, 07:54:02  
Job time : 29 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 24, 2005, 07:49:26 ; Search time 81 Seconds  
(without alignments)  
2511.184 Million cell updates/sec

Title: US-09-936-759-6

Perfect score: 3001

Sequence: 1 MVRPQRNKKRIFILLNGVWN.....TRDRQPKLVAVHLRLMSEV 563

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1608061 seqs, 361289386 residues

Total number of hits satisfying chosen parameters: 1608061

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep:\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep:\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUBCOMB.pep:\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep:\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep:\*
- 6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep:\*
- 7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep:\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep:\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep:\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep:\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep:\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep:\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep:\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep:\*
- 15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep:\*
- 16: /cgn2\_6/ptodata/2/pubpaa/US10D\_PUBCOMB.pep:\*
- 17: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep:\*
- 18: /cgn2\_6/ptodata/2/pubpaa/US11\_NEW\_PUB.pep:\*
- 19: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep:\*
- 20: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2999	99.9	563	US-10-364-649-6	Sequence 6, Appli
2	2999	99.9	563	US-10-364-649-21	Sequence 21, Appli
3	2999	99.9	563	US-10-369-493-2995	Sequence 2995, Ap
4	929	31.0	602	US-10-120-145-2	Sequence 2, Appli
5	929	31.0	602	US-10-364-649-2	Sequence 8, Appli
6	929	31.0	602	US-10-364-649-8	Sequence 2, Appli
7	929	31.0	602	US-10-364-649-15	Sequence 15, Appli
8	929	31.0	618	US-10-120-145-4	Sequence 4, Appli
9	925	30.8	607	US-10-120-145-8	Sequence 8, Appli
10	925	30.8	607	US-10-364-649-18	Sequence 18, Appli
11	925	30.8	615	US-10-364-649-28	Sequence 28, Appli
12	919.5	30.6	598	US-09-862-660-2	Sequence 2, Appli
13	919.5	30.6	598	US-10-673-935-2	Sequence 2, Appli

14	899.5	30.0	648	14	US-10-421-175-4	Sequence 4, Appli
15	882	29.4	1010	9	US-09-118-276-12	Sequence 12, Appli
16	882	29.4	1010	16	US-10-705-197A-12	Sequence 12, Appli
17	879	29.3	603	14	US-10-161-403-106	Sequence 106, App
18	879	29.3	603	14	US-10-120-145-6	Sequence 6, Appli
19	879	29.3	603	14	US-10-364-649-17	Sequence 17, Appli
20	879	29.3	603	14	US-10-364-649-23	Sequence 23, Appli
21	879	29.3	603	14	US-10-369-493-842	Sequence 842, App
22	879	29.3	603	17	US-10-161-408-17	Sequence 17, Appli
23	877	29.2	618	14	US-10-356-088-28	Sequence 28, Appli
24	877	29.2	618	16	US-10-799-326-28	Sequence 28, Appli
25	875	29.2	604	10	US-09-893-525-37	Sequence 37, Appli
26	875	29.2	604	17	US-10-763-380-37	Sequence 37, Appli
27	875	29.2	659	10	US-09-893-525-40	Sequence 40, Appli
28	875	29.2	659	17	US-10-763-380-40	Sequence 40, Appli
29	875	29.2	850	10	US-09-893-525-42	Sequence 42, Appli
30	875	29.2	850	17	US-10-763-380-42	Sequence 42, Appli
31	872.5	29.1	602	13	US-10-195-158-5	Sequence 5, Appli
32	872.5	29.1	602	14	US-10-195-518-5	Sequence 5, Appli
33	872.5	29.1	607	14	US-10-369-493-15291	Sequence 15291, A
34	872.5	29.1	711	14	US-10-338-411-23	Sequence 23, Appli
35	872.5	29.1	711	15	US-10-389-640-23	Sequence 23, Appli
36	865.5	28.8	613	14	US-10-120-145-5	Sequence 5, Appli
37	865.5	28.8	613	14	US-10-364-649-16	Sequence 16, Appli
38	865.5	28.8	651	14	US-10-421-175-2	Sequence 2, Appli
39	865.5	28.8	722	14	US-10-136-841-6	Sequence 6, Appli
40	865.5	28.8	722	15	US-10-272-531A-6	Sequence 6, Appli
41	865.5	28.8	722	15	US-10-272-483A-6	Sequence 6, Appli
42	705.5	23.5	450	14	US-10-356-088-40	Sequence 40, Appli
43	705.5	23.5	450	16	US-10-799-326-40	Sequence 40, Appli
44	692.5	23.1	376	14	US-10-364-649-4	Sequence 4, Appli
45	692.5	23.1	376	14	US-10-364-649-19	Sequence 19, Appli

ALIGNMENTS

RESULT 1  
US-10-364-649-6  
; Sequence 6, Application US/10364649  
; Publication No. US20030229921A1  
; GENERAL INFORMATION:  
; APPLICANT: Richard A. Jefferson and Jorge E. Mayer  
; TITLE OF INVENTION: MICROBIAL B-GLUCURONIDASE GENES, GENE  
; TITLE OF INVENTION: PRODUCTS, AND USES THEREOF  
; FILE REFERENCE: 190106.405C1  
; CURRENT APPLICATION NUMBER: US/10/364, 649  
; CURRENT FILING DATE: 2003-02-12  
; PRIOR APPLICATION NUMBER: 10/364, 649  
; PRIOR FILING DATE: 2003-02-12  
; PRIOR APPLICATION NUMBER: US 09/270, 957  
; PRIOR FILING DATE: 1999-03-17  
; NUMBER OF SEQ ID NOS: 112  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 6  
; LENGTH: 563  
; TYPE: PRT  
; ORGANISM: Thermotoga maritima  
; FEATURE:  
; NAME/KEY: VARIANT  
; LOCATION: (1)...(563)  
; OTHER INFORMATION: Xaa = Any Amino Acid  
US-10-364-649-6  
Query Match 99.9%; Score 2999, DB 14; length 563;  
Best Local Similarity 100.0%; Pred. No. 1.8e-241;  
Matches 563; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MVRPQRNKKRIFILLNGVWNLEVTSKDRPIAVPGSWNEQYODLCYEEGPFYKTFYVPK 60  
Db 1 MVRPQRNKKRIFILLNGVWNLEVTSKDRPIAVPGSWNEQYODLCYEEGPFYKTFYVPK 60  
QY 61 XLSQKHIRLYFAAVNTDCEVFLNGEKVGENHIEYLPFEVDVTGKVKSGENELRVVVENRL 120



```
DB 61 XLSQKHIRLYFAAVNTDCEVFLNGEKVGENHIEYLPFEVDVTGKVKSGENELRVVVENRL 120
QY 121 KVGFPSPKVPDSGTHTVGFFGSPFANPDFFPYGGIIRPVLIETDTHARILDIWVDTSES 180
DB 121 KVGFPSPKVPDSGTHTVGFFGSPFANPDFFPYGGIIRPVLIETDTHARILDIWVDTSES 180
QY 181 EPEKLGKVKVKIEVSEEAAGVQEMTIKLGEEBKIRTSNRFVEGEFILLENAFWSLEDPY 240
DB 181 EPEKLGKVKVKIEVSEEAAGVQEMTIKLGEEBKIRTSNRFVEGEFILLENAFWSLEDPY 240
QY 241 LYPLKVELKEDEYTLDIGIRTIISWDEKRLYLNGKPVFLKGFGKHEEPVLGOGTFYPLMI 300
DB 241 LYPLKVELKEDEYTLDIGIRTIISWDEKRLYLNGKPVFLKGFGKHEEPVLGOGTFYPLMI 300
QY 301 KDFNLLKWINANSFRTSHYPYSEEWLADRLGILVIDEAPHVGITRHYNPETOKIAED 360
DB 301 KDFNLLKWINANSFRTSHYPYSEEWLADRLGILVIDEAPHVGITRHYNPETOKIAED 360
QY 361 NIRRMIDRHKNHPSVIMSVANEPESNHPDAEGFFKALYETANEMDRTRPVVWVSMMDAP 420
DB 361 NIRRMIDRHKNHPSVIMSVANEPESNHPDAEGFFKALYETANEMDRTRPVVWVSMMDAP 420
QY 421 DERTRDVALKYFDIVCVNRYGYWYIYQGRIEEGLQALEKDIIEELYARHRKPIFVTEFGAD 480
DB 421 DERTRDVALKYFDIVCVNRYGYWYIYQGRIEEGLQALEKDIIEELYARHRKPIFVTEFGAD 480
QY 481 AIAGIHYPDPQMFSEBYQAEIYEKTIIRLLKKDYIIGTHWAFADFKTPQNVRRPILNHK 540
DB 481 AIAGIHYPDPQMFSEBYQAEIYEKTIIRLLKKDYIIGTHWAFADFKTPQNVRRPILNHK 540
QY 541 GVFTTRDRQPKLVAVHLRRLMSEV 563
DB 541 GVFTTRDRQPKLVAVHLRRLMSEV 563
```

## RESULT 2

```
US-10-364-649-21
; Sequence 21, Application US/10364649
; Publication No. US20030229921A1
; GENERAL INFORMATION:
; APPLICANT: Richard A. Jefferson and Jorge E. Mayer
; TITLE OF INVENTION: MICROBIAL B-GLUCURONIDASE GENES, GENE
; FILE REFERENCE: 190106.405C1
; CURRENT APPLICATION NUMBER: US/10/364,649
; PRIOR FILING DATE: 2003-02-12
; PRIOR APPLICATION NUMBER: 10/364,649
; PRIOR FILING DATE: 2003-02-12
; PRIOR APPLICATION NUMBER: US 09/270,957
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21
; LENGTH: 563
; TYPE: PRT
; ORGANISM: Thermotoga maritima
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(563)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-10-364-649-21
```

Query Match 99.9%; Score 2999; DB 14; Length 563;  
Best Local Similarity 100.0%; Pred. No. 1.8e-241;  
Matches 563; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 MVRPQNRKRFILILNGVNLVETSKDRPIAVPGSWNEQYQDLCYEEGPFTYKTTFYVPK 60
DB 1 MVRPQNRKRFILILNGVNLVETSKDRPIAVPGSWNEQYQDLCYEEGPFTYKTTFYVPK 60
QY 61 XLSQKHIRLYFAAVNTDCEVFLNGEKVGENHIEYLPFEVDVTGKVKSGENELRVVVENRL 120
DB 61 XLSQKHIRLYFAAVNTDCEVFLNGEKVGENHIEYLPFEVDVTGKVKSGENELRVVVENRL 120
```

```
DB 61 XLSQKHIRLYFAAVNTDCEVFLNGEKVGENHIEYLPFEVDVTGKVKSGENELRVVVENRL 120
QY 121 KVGFPSPKVPDSGTHTVGFFGSPFANPDFFPYGGIIRPVLIETDTHARILDIWVDTSES 180
DB 121 KVGFPSPKVPDSGTHTVGFFGSPFANPDFFPYGGIIRPVLIETDTHARILDIWVDTSES 180
QY 181 EPEKLGKVKVKIEVSEEAAGVQEMTIKLGEEBKIRTSNRFVEGEFILLENAFWSLEDPY 240
DB 181 EPEKLGKVKVKIEVSEEAAGVQEMTIKLGEEBKIRTSNRFVEGEFILLENAFWSLEDPY 240
QY 241 LYPLKVELKEDEYTLDIGIRTIISWDEKRLYLNGKPVFLKGFGKHEEPVLGOGTFYPLMI 300
DB 241 LYPLKVELKEDEYTLDIGIRTIISWDEKRLYLNGKPVFLKGFGKHEEPVLGOGTFYPLMI 300
QY 301 KDFNLLKWINANSFRTSHYPYSEEWLADRLGILVIDEAPHVGITRHYNPETOKIAED 360
DB 301 KDFNLLKWINANSFRTSHYPYSEEWLADRLGILVIDEAPHVGITRHYNPETOKIAED 360
QY 361 NIRRMIDRHKNHPSVIMSVANEPESNHPDAEGFFKALYETANEMDRTRPVVWVSMMDAP 420
DB 361 NIRRMIDRHKNHPSVIMSVANEPESNHPDAEGFFKALYETANEMDRTRPVVWVSMMDAP 420
QY 421 DERTRDVALKYFDIVCVNRYGYWYIYQGRIEEGLQALEKDIIEELYARHRKPIFVTEFGAD 480
DB 421 DERTRDVALKYFDIVCVNRYGYWYIYQGRIEEGLQALEKDIIEELYARHRKPIFVTEFGAD 480
QY 481 AIAGIHYPDPQMFSEBYQAEIYEKTIIRLLKKDYIIGTHWAFADFKTPQNVRRPILNHK 540
DB 481 AIAGIHYPDPQMFSEBYQAEIYEKTIIRLLKKDYIIGTHWAFADFKTPQNVRRPILNHK 540
QY 541 GVFTTRDRQPKLVAVHLRRLMSEV 563
DB 541 GVFTTRDRQPKLVAVHLRRLMSEV 563
```

## RESULT 3

```
US-10-369-493-2995
; Sequence 2995, Application US/10369493
; Publication No. US2003023675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; NUMBER OF SEQ ID NOS: 2002-02-21
; SEQ ID NO 2995
; LENGTH: 563
; TYPE: PRT
; ORGANISM: Thermotoga maritima
US-10-369-493-2995
```

Query Match 99.9%; Score 2999; DB 14; Length 563;  
Best Local Similarity 99.8%; Pred. No. 1.8e-241;  
Matches 562; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
QY 1 MVRPQNRKRFILILNGVNLVETSKDRPIAVPGSWNEQYQDLCYEEGPFTYKTTFYVPK 60
DB 1 MVRPQNRKRFILILNGVNLVETSKDRPIAVPGSWNEQYQDLCYEEGPFTYKTTFYVPK 60
QY 61 XLSQKHIRLYFAAVNTDCEVFLNGEKVGENHIEYLPFEVDVTGKVKSGENELRVVVENRL 120
DB 61 XLSQKHIRLYFAAVNTDCEVFLNGEKVGENHIEYLPFEVDVTGKVKSGENELRVVVENRL 120
QY 121 KVGFPSPKVPDSGTHTVGFFGSPFANPDFFPYGGIIRPVLIETDTHARILDIWVDTSES 180
DB 121 KVGFPSPKVPDSGTHTVGFFGSPFANPDFFPYGGIIRPVLIETDTHARILDIWVDTSES 180
```



Db 121 KVGFPSPKVPDSGTHTVGFGSGPPANFDFFPYGIIIRPVLIETPDHARILDIWDTSES 180  
Qy 181 EPEKTLGKVKVIEVSEAVGQEMTIKLGEEKKIRTSNRFVGEFIIENARFWSLEDPY 240  
Db 181 EPEKTLGKVKVIEVSEAVGQEMTIKLGEEKKIRTSNRFVGEFIIENARFWSLEDPY 240  
Qy 241 LYPKVELEKDEYTLDIGIRTSWDEKRLYLNGKVPFLKFGKHEEPVLGGTFYPLMI 300  
Db 241 LYPKVELEKDEYTLDIGIRTSWDEKRLYLNGKVPFLKFGKHEEPVLGGTFYPLMI 300  
Qy 301 KDFNLKWINANSFRTSHYPYSEEWLDLADRLGILVIDEAPHVGITRHYHNPETOKIAED 360  
Db 301 KDFNLKWINANSFRTSHYPYSEEWLDLADRLGILVIDEAPHVGITRHYHNPETOKIAED 360  
Qy 361 NIRRMIDRHKHPSVIMSVANPEESNHPDAEGFFKALYETANEMDRTRPVVMVSMMDAP 420  
Db 361 NIRRMIDRHKHPSVIMSVANPEESNHPDAEGFFKALYETANEMDRTRPVVMVSMMDAP 420  
Qy 421 DERTRDVALKYPDIVCVNRYGYWYIYQGRIEEGLQALEKDIIEELYARHRKPIFVTEFGAD 480  
Db 421 DERTRDVALKYPDIVCVNRYGYWYIYQGRIEEGLQALEKDIIEELYARHRKPIFVTEFGAD 480  
Qy 481 AIAGIHYDPQMFSEEQAEVLEKTIIRLLKKDYIIGTHWAFADFKTPQNVRRPILNHK 540  
Db 481 AIAGIHYDPQMFSEEQAEVLEKTIIRLLKKDYIIGTHWAFADFKTPQNVRRPILNHK 540  
Qy 541 GVFTDRQPKLVAAHLRLMSEV 563  
Db 541 GVFTDRQPKLVAAHLRLMSEV 563

RESULT 4  
US-10-120-145-2

; Sequence 2, Application US/10120145  
; Publication No. US20030157684A1  
; GENERAL INFORMATION:  
; APPLICANT: Jefferson, Richard A.  
; APPLICANT: Kilian, Andrzej  
; APPLICANT: Keese, Paul Konrad  
; TITLE OF INVENTION: MICROBIAL BETA-GLUCURONIDASE GENES, GENE PRODUCTS AND  
; FILE REFERENCE: 190106.405  
; CURRENT APPLICATION NUMBER: US/10/120,145  
; CURRENT FILING DATE: 2002-04-11  
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/149,727  
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-09-08  
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 60/058,263  
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-09-09  
; NUMBER OF SEQ ID NOS: 71  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 602  
; TYPE: PRT  
; ORGANISM: Bacillus sp.  
US-10-120-145-2

Query Match 31.0%; Score 929; DB 14; Length 602;  
Best Local Similarity 35.6%; Pred. No. 1.5e-68;  
Matches 221; Conservative 93; Mismatches 224; Indels 82; Gaps 15;

Qy 1 MVRPQRNKKRFFILILNGVWNLV-----TSKDRPIAVPGSWNE--QYQDLC 44  
Db 1 MLYPINTETRGVFDLNGVWNLKDYGKLEEKWYESKLTDTTISMAVPSSYNDIGVTKEIR 60  
Qy 45 YEEGPFTYKTFYVPKLSQKHRLRYFAAVNTDCEVFLNGEKGVENHIEYLPREVDTGK 104  
Db 61 NHIGVWYEREFVTPAYLKQDRIVLRFGSATHKAIYVYNGELVVEHKGFLPEAEINNS 120  
Qy 105 VKSGENELRVVVENRLKVGFPSPKVPDSGTHTVGFGS-----FPPANDFPFPYG 154  
Db 121 LRDGMNRVTVAVDNIL-----DDSTLPVGLYSERHEEGLGKIVIRNKPNDFFFNYA 170  
Qy 155 GIIRPVLIEFTDHARILDIWDTSESEPEKKLGKVKVIEVSEAVGQEMTIKLG--EEE 212

Db 171 GLHRPVKIYTFPFTYVEDISVTDENGP---TGTVTYTVDFQ-----GKAETVKSVVDEB 223  
Qy 213 KKI RTSNRFVEGEFIIENARFWSLEDPYLPLKYLEKDEYTLDI-----GIRTSWDEK 267  
Db 224 GKVAVSTEGLSGVAEIPNVILMEPLNTYLYQIKVELVNDGLTIDVYEEFPGVRTVEVNDG 283  
Qy 268 RLYLNGKVPFLKFGKHEEPVLGGTFYPLMIKDFNLKWINANSFRTSHYPYSEEWLD 327  
Db 284 KFLINNKPFYFRGFGKHEDTPINGRGFNEASVMDFNILKWINANSFRTAHYPSSELMR 343  
Qy 328 LADRLGILVIDEAPHVGITRHYN-----PETOKIA-----EDNIRRMIDR 368  
Db 344 LADREGLVVIDETPAVG--HLMFMATYGLGEGSERVSTWEKIRTFEHQDVLRELVS 400  
Qy 369 HKNHPSVIMSVANPEESNHPDAEGFFKALYETANEMD-RTRPVVMVSMMDAPDERTDV 427  
Db 401 DKNHPSVIMSVANPEESNHPDAEGFFKALYETANEMD-RTRPVVMVSMMDAPDERTDV 427  
Qy 428 ALKYPDIVCVNRYGYWYIYQGRIEEGLQALEKDIIEELYARHR---KPIFVTEFGADAIA 483  
Db 459 VAEIDVIALNRYNGWYFPDGDLEAAKVHLR---QEFHAMNKRCPGKPIMITETEGADTVA 515  
Qy 484 GIHYDPQMFSEEQAEVLEKTIIRLLKKDYIIGTHWAFADFKTPQNVRRPILNHKGVF 543  
Db 516 GFHDIDPVMFTBEYQVEYQANHVVFDEFENFVGEQAMNFADPAITSQGVMRVQGNKKGVF 575  
Qy 544 TRDRQPKLVAAHLRLMSEV 563  
Db 576 TRDRKPKLAAHVFRERWTNI 595

RESULT 5  
US-10-364-649-2

; Sequence 2, Application US/10364649  
; Publication No. US20030229921A1  
; GENERAL INFORMATION:  
; APPLICANT: Richard A. Jefferson and Jorge E. Mayer  
; TITLE OF INVENTION: MICROBIAL B-GLUCURONIDASE GENES, GENE  
; FILE REFERENCE: 190106.405C1  
; CURRENT APPLICATION NUMBER: US/10/364,649  
; CURRENT FILING DATE: 2003-02-12  
; PRIOR APPLICATION NUMBER: 10/364,649  
; PRIOR FILING DATE: 2003-02-12  
; PRIOR APPLICATION NUMBER: US 09/270,957  
; PRIOR FILING DATE: 1999-03-17  
; NUMBER OF SEQ ID NOS: 112  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 602  
; TYPE: PRT  
; ORGANISM: Bacillus sp.  
US-10-364-649-2

Query Match 31.0%; Score 929; DB 14; Length 602;  
Best Local Similarity 35.6%; Pred. No. 1.5e-68;  
Matches 221; Conservative 93; Mismatches 224; Indels 82; Gaps 15;

Qy 1 MVRPQRNKKRFFILILNGVWNLV-----TSKDRPIAVPGSWNE--QYQDLC 44  
Db 1 MLYPINTETRGVFDLNGVWNLKDYGKLEEKWYESKLTDTTISMAVPSSYNDIGVTKEIR 60  
Qy 45 YEEGPFTYKTFYVPKLSQKHRLRYFAAVNTDCEVFLNGEKGVENHIEYLPREVDTGK 104  
Db 61 NHIGVWYEREFVTPAYLKQDRIVLRFGSATHKAIYVYNGELVVEHKGFLPEAEINNS 120  
Qy 105 VKSGENELRVVVENRLKVGFPSPKVPDSGTHTVGFGS-----FPPANDFPFPYG 154  
Db 121 LRDGMNRVTVAVDNIL-----DDSTLPVGLYSERHEEGLGKIVIRNKPNDFFFNYA 170  
Qy 155 GIIRPVLIEFTDHARILDIWDTSESEPEKKLGKVKVIEVSEAVGQEMTIKLG--EEE 212

Db 171 GLHRPVKIYTTPTTYVEDISVTDENGP---TGTVTYTVDFQ-----GKAETVKVSVVDEE 223  
QY 213 KKIITSNRFEVEGEFIELENARFWSLEDPLYLPLKVELEKDEYTLDI-----GIRTSWDEK 267  
Db 224 GKVAASTEGLSGNVEIPNVILMEPLNTYLYQIKVELVNDGLTIDVYEEPFQVTRVEVNDG 283  
QY 268 RLYLNGKPVFLKGFKGHEEFVLGGQTFYPLMIKDFNLLKMINANSFRTSHYPYSEEWLD 327  
Db 284 KFLINNKPFYFGFGKHEDTPINGRGFNEASNVMDFNILKMWIGANSFRTAHYPSSEELMR 343  
QY 328 LADRLGILVIDEAPHVGTIRYHYN-----PETOKIA-----EDNIRRMIDR 368  
Db 344 LADREGLVVIDETPAVG---HLNFMATTGLGEGSERVSTWEKIRTFEHQDVLRELVS 400  
QY 369 HKNHPSVIMSVANEPESNHPDAEGFFKALYETANEMD-RTRPVVMVSMMDAPDERTRDV 427  
Db 401 DKNHPSVVMWSIANEAATEEGAYEYFKPLVELTKELDPQKRPTIIVLFVMAFPE--TDK 458  
QY 428 ALKYPDVCVNRYYGWIYQGRTEEGQLALEKDIIEELYARHR----KPIFVTEFGADAIA 483  
Db 459 VAEILDVIALNRNGWYFDGGDLBAKVHLR---QEFHAMNKRCPGKPMITEYGADTV 515  
QY 484 GIHYDPQMFSEYQAEVLEKTRILLKKDYIIGTHVMAFADFKTPQNVRRPILNHKGVF 543  
Db 516 GFHDIDPVMFTEEYQVEYYQANHVFDEFENFVGEQAMNFADFATSQGVMRVQGNKKGVF 575  
QY 544 TRDRQPKLVAVHLRLMSEV 563  
Db 576 TRDRKPKLAAHVFRERWTNI 595

RESULT 6  
US-10-364-649-8

; Sequence 8, Application US/10364649  
; Publication No. US20030229921A1  
; GENERAL INFORMATION:  
; APPLICANT: Richard A. Jefferson and Jorge E. Mayer  
; TITLE OF INVENTION: MICROBIAL B-GLUCURONIDASE GENES, GENE  
; TITLE OF INVENTION: PRODUCTS, AND USES THEREOF  
; FILE REFERENCE: 190106.405C1  
; CURRENT APPLICATION NUMBER: US/10/364,649  
; CURRENT FILING DATE: 2003-02-12  
; PRIOR APPLICATION NUMBER: 10/364,649  
; PRIOR FILING DATE: 2003-02-12  
; PRIOR APPLICATION NUMBER: US 09/270,957  
; PRIOR FILING DATE: 1999-03-17  
; NUMBER OF SEQ ID NOS: 112  
; SOFTWARE: FastSeq for windows Version 4.0  
; SEQ ID NO 8  
; LENGTH: 602  
; TYPE: PRT  
; ORGANISM: Bacillus sp.  
; US-10-364-649-8

Query Match 31.0%; Score 929; DB 14; Length 602;  
Best Local Similarity 35.6%; Pred. No. 1.5e-68;  
Matches 221; Conservative 93; Mismatches 224; Indels 82; Gaps 15;

QY 1 MVRPQRNKKRIFILLNGVMNLEV-----TSKDRPIAVPGSWNE--QYODLC 44  
Db 1 MLYPINTETRGVFDLNGVMNFKLDYKGLBEKWEYSKLTDTISMAVPSYNDIGVTKEIR 60  
QY 45 YEEGPFTYKTTFFYPKXLSQKHIRLYFAAVNTDCEVFLNGEKGVENHIEVLPFEVDVTGK 104  
Db 61 NHIGYVWEREFVTPAYLKDQRIVLRFGSATHKAIYVNGELVVEHKGFLPFEAEINNS 120  
QY 105 VKSGENELRVVENRLKVGFPSPKVPDSGTHTVGFFGS-----FPPANFDFFPYG 154  
Db 121 LRDGMNRTVAVDNLL-----DSTLPVGLYSEHHEGLGKVI RNKPNFDFFNVA 170  
QY 155 GIIRPVLIIEFTDHARILDIWVDTSESEPEKKLGKVKIIVSEEA VGOEMTIKLG--EEB 212  
Db 171 GLHRPVKIYTTPTTYVEDISVTDENGP---TGTVTYTVDFQ-----GKAETVKVSVVDEE 223

QY 213 KKIITSNRFEVEGEFIELENARFWSLEDPLYLPLKVELEKDEYTLDI-----GIRTSWDEK 267  
Db 224 GKVAASTEGLSGNVEIPNVILMEPLNTYLYQIKVELVNDGLTIDVYEEPFQVTRVEVNDG 283  
QY 268 RLYLNGKPVFLKGFKGHEEFVLGGQTFYPLMIKDFNLLKMINANSFRTSHYPYSEEWLD 327  
Db 284 KFLINNKPFYFGFGKHEDTPINGRGFNEASNVMDFNILKMWIGANSFRTAHYPSSEELMR 343  
QY 328 LADRLGILVIDEAPHVGTIRYHYN-----PETOKIA-----EDNIRRMIDR 368  
Db 344 LADREGLVVIDETPAVG---HLNFMATTGLGEGSERVSTWEKIRTFEHQDVLRELVS 400  
QY 369 HKNHPSVIMSVANEPESNHPDAEGFFKALYETANEMD-RTRPVVMVSMMDAPDERTRDV 427  
Db 401 DKNHPSVVMWSIANEAATEEGAYEYFKPLVELTKELDPQKRPTIIVLFVMAFPE--TDK 458  
QY 428 ALKYPDVCVNRYYGWIYQGRTEEGQLALEKDIIEELYARHR----KPIFVTEFGADAIA 483  
Db 459 VAEILDVIALNRNGWYFDGGDLBAKVHLR---QEFHAMNKRCPGKPMITEYGADTV 515  
QY 484 GIHYDPQMFSEYQAEVLEKTRILLKKDYIIGTHVMAFADFKTPQNVRRPILNHKGVF 543  
Db 516 GFHDIDPVMFTEEYQVEYYQANHVFDEFENFVGEQAMNFADFATSQGVMRVQGNKKGVF 575  
QY 544 TRDRQPKLVAVHLRLMSEV 563  
Db 576 TRDRKPKLAAHVFRERWTNI 595

RESULT 7  
US-10-364-649-15

; Sequence 15, Application US/10364649  
; Publication No. US20030229921A1  
; GENERAL INFORMATION:  
; APPLICANT: Richard A. Jefferson and Jorge E. Mayer  
; TITLE OF INVENTION: MICROBIAL B-GLUCURONIDASE GENES, GENE  
; TITLE OF INVENTION: PRODUCTS, AND USES THEREOF  
; FILE REFERENCE: 190106.405C1  
; CURRENT APPLICATION NUMBER: US/10/364,649  
; CURRENT FILING DATE: 2003-02-12  
; PRIOR APPLICATION NUMBER: 10/364,649  
; PRIOR FILING DATE: 2003-02-12  
; PRIOR APPLICATION NUMBER: US 09/270,957  
; PRIOR FILING DATE: 1999-03-17  
; NUMBER OF SEQ ID NOS: 112  
; SOFTWARE: FastSeq for windows Version 4.0  
; SEQ ID NO 15  
; LENGTH: 602  
; TYPE: PRT  
; ORGANISM: Bacillus sp.  
; US-10-364-649-15

Query Match 31.0%; Score 929; DB 14; Length 602;  
Best Local Similarity 35.6%; Pred. No. 1.5e-68;  
Matches 221; Conservative 93; Mismatches 224; Indels 82; Gaps 15;

QY 1 MVRPQRNKKRIFILLNGVMNLEV-----TSKDRPIAVPGSWNE--QYODLC 44  
Db 1 MLYPINTETRGVFDLNGVMNFKLDYKGLBEKWEYSKLTDTISMAVPSYNDIGVTKEIR 60  
QY 45 YEEGPFTYKTTFFYPKXLSQKHIRLYFAAVNTDCEVFLNGEKGVENHIEVLPFEVDVTGK 104  
Db 61 NHIGYVWEREFVTPAYLKDQRIVLRFGSATHKAIYVNGELVVEHKGFLPFEAEINNS 120  
QY 105 VKSGENELRVVENRLKVGFPSPKVPDSGTHTVGFFGS-----FPPANFDFFPYG 154  
Db 121 LRDGMNRTVAVDNLL-----DSTLPVGLYSEHHEGLGKVI RNKPNFDFFNVA 170  
QY 155 GIIRPVLIIEFTDHARILDIWVDTSESEPEKKLGKVKIIVSEEA VGOEMTIKLG--EEB 212  
Db 171 GLHRPVKIYTTPTTYVEDISVTDENGP---TGTVTYTVDFQ-----GKAETVKVSVVDEE 223

[illegible]

## RESULT 8

US-10-120-145-4  
 ; Sequence 4, Application US/10120145  
 ; Publication No. US20030157684A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Jefferson, Richard A.  
 ; APPLICANT: Kilian, Andrzej  
 ; APPLICANT: Keese, Paul Konrad  
 ; TITLE OF INVENTION: MICROBIAL BETA-GLUCURONIDASE GENES, GENE PRODUCTS AND  
 ; TITLE OF INVENTION: USES THEREOF  
 ; FILE REFERENCE: 190106.405  
 ; CURRENT APPLICATION NUMBER: US/10/120,145  
 ; CURRENT FILING DATE: 2002-04-11  
 ; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/149,727  
 ; PRIOR FILING DATE: EARLIER FILING DATE: 1998-09-08  
 ; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 60/058,263  
 ; PRIOR FILING DATE: EARLIER FILING DATE: 1997-09-09  
 ; NUMBER OF SEQ ID NOS: 71  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 4  
 ; LENGTH: 618  
 ; TYPE: PRT  
 ; ORGANISM: Bacillus sp.  
 US-10-120-145-4

## Query Match

Query Match	31.0%;	Score 929;	DB 14;	Length 618;
Best Local Similarity	35.6%;	Pred. No. 1.6e+68;		
Matches 221;	Conservative 93;	Mismatches 224;	Indels 82;	Gaps 15;

```

QY      1 MVRPORNKKRFILLNGVMNLEV-----TSKDRPIAVEGSWNE--QYODLC 44
      | : | : | | | | | : : | : | | : | : : :
Db      17 MLYPINTETRGVFDLNGVMNFKLDYGKGLSEKMYESKLTDTISMAVPSYNDIGVYKEIR 76

QY      45 YEEGPFYKTFYVPKKLSQKHRLYPAAVNTDCEVFLNGEKGXGENTHIEYLPFEVDVTGK 104
      | : | : | | | | : : | : | | : | : : :
Db      77 NHIGYVMYEREFYVPAYLKQDRIVLRFGSATHKALYYVNGELVVEHKGGLPFPEAEIINS 136

QY      105 VKSGENELRVVVENRLKVGFPKVPDSGTHTVGFGS-----FPRANDFPPYG 154
      | : | : | : | : | : | : | : | : | : | :
Db      137 LRDMNRVTVAVDNIL-----DSTLPVGLYSERHEGLGKVI RNKPNDFFNYA 186

QY      155 GIIRPVLIEFTDHARILDIMWDTSESEPEKLGKVKKIEVSEEAAGCEMTIKG--EE 212
      | : | | | : | | | : | : | : : : : : :
Db      187 GJHRPVKIYTTPEYVEDISVTDENGPR--TGIVYTVDFQ---GAETVKVSVDDE 239

```

QY	213	KIIRTSNRPFVEGEFIIENARFWSJEDPYLPYLKYLEEKEDEYTLDI-----GIRTI5WDEK	267
Db	240	GKVASTEGLSGNVEIPNVILWEPIINTYLYQIKVELVNDGLTIDVVEEPFGRVTEVNDG	299
QY	268	RLVYLNKQPVFLKGFGEKHEEPVJQGTFFPLMKIDENLLKMINANSFRTSHYPYSEEWLD	327
Db	300	KELINNKPFYFKGFGKHEDTPINGRGFNEASNVDENILKMWIGANSFRTAHPYSEELMR	359
QY	328	LADRLGILVIDEAPHVIGITRYHN-----PETOKIA-----EDNIRRMIDR	368
Db	360	LADREGLVVIDETPAVG--HINFMATTGLGEGSERVSTWEKIRTFEHHQDVLRELYSR	416
QY	369	HKNHPSVIMWSVANEPESNHPDAEGFKALYETANEMD-RTRPVVWVSMMDAPDERTDY	427
Db	417	DKNHPSVVMWSIANEATEEBEGAYEYFKPLVELTRELDPQKRPTVIVLFVMATPE--TDK	474
QY	428	ALKYFDIVCVNRYGWIYYQGRTEBGLQALEKDIIEELYARHR---KPIFVTEFGADAIA	483
Db	475	VAEIIDVIALNRNGWYFDDGDLLEAKVHLR---QEFHAMNKRCPGKPMITEYGADTVA	531
QY	484	GIHYDPQPMFSEFYQALVEKTRLLKKDYITGTHVWAFADFKTPQNVRRPILNHKGVF	543
Db	532	GPHDIDPVMFTEEYQVEYYQANHVVFDEFENFVEQAMNFADFATSQGVMRVQGNKKGVF	591
QY	544	TRDRQPKLVAAVHLRLMSEV	563
Db	592	TRDRKPKLAAHVFRERWTNI	611

## RESULT 9

```

US-10-120-145-8
; Sequence 8, Application US/10120145
; Publication No. US20030157684A1
; GENERAL INFORMATION:
; APPLICANT: Jefferson, Richard A.
; APPLICANT: Kilian, Andrzej
; APPLICANT: Keese, Paul Konrad
; TITLE OF INVENTION: MICROBIAL BETA-GLUCURONIDASE GENES, GENE PRODUCTS AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: 190106.405
; CURRENT APPLICATION NUMBER: US/10/120,145
; CURRENT FILING DATE: 2002-04-11
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/149,727
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-09-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 60/058,263
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-09-09
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 607
; TYPE: PRT
; ORGANISM: Bacillus sp.
US-10-120-145-8

```

## Query Match

Query Match	30.8%;	Score 925;	DB 14;	Length 607;
Best Local Similarity	35.7%;	Pred. No. 3.3e-68;		
Matches 220;	Conservative 92;	Mismatches 223;	Indels 82;	Gaps 15;

```

QY      4 PQRNKRFFILNGVMNLEV-----TSKDRPIAVPGSMNE--QYODLCYEE 47
      |  :  :  |||||  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      9 PINTETRGVEDLNGVMNPKLDYGGKLEEKWYESKCLDTISMAVPSSTYNDIGVTKEIRNHI 68
QY      48 GPFTYKTTFFVPEKKSQKHIRLYFAAVNTDCEVFLNGEKGVENHIEFLPFEDVTGKYKS 107
      |  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      69 GYVMYEREFVPAYLKDQRIVIRFGSATHKALVYVNGELVVEHKGGFLPFEEAIRNNSLD 128
QY      108 GENELRVVVENRLKVGGFPSKVPDSGTHTVGFGS-----PPANFDFFPYGII 157
      |  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      129 GMRNRTVAVDNILL-----DDSTLPVGLYSERHBEGLGKVIIRNKNPFDFFNYAGLH 178
QY      158 RPLVLIETDHRILDIWDTSESEPEKKLGKVKYKIEVSEAVGQEMTIKLG--EEEKKI 215
      |  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :

```

```
Db 179 RPKVIYTPFTYVEDISVTDENGPR--TGTVTYTVDFQ---GKAETVKVSVVDEEGKV 231
QY 216 RTSNRFVEGEFILENARFWSLEDPLYPLKVELEKDEYTLDI-----GIRTISWDEKRLY 270
Db 232 VASTEGISGNVEIPNVILMEPLNTLYQIKVELVNDGLTIDVYEEBPGVRTVEVNDGKFL 291
QY 271 LNKGPVFLKGFKGHEEPVLGQGTFFPLMIKD FNLLKWINANSFRTSHYPYSEEWLAD 330
Db 292 INNKPFFYFKGFKGHEEDTPINGRGFNBSNVMDFNILKWINANSFRTAHYPYSEELMLAD 351
QY 331 RLGLIVIDEAPHVGITRHYHN-----PETQKIA-----EDNIRRMIDRHKN 371
Db 352 REGLVVIDETPAVG--HINFMATTGLGEGSERVSTWEKIRTFEHHQDVLRELVSBDKN 408
QY 372 HPSVIMSVANEPESNHPDAEGFFKALYETANEMD-RTRPVVMVSMMDAPDERTRDVALK 430
Db 409 HPSVIMWSIANEAATEEGAYEYFKPLVELTKELDPQKRPVTIVLFVWATPE--TDKVAE 466
QY 431 YFDIVCNRYYGWYIYQGRIEEGLQALEKDI EELYARHR---KPIFVTEFGADAIAGIH 486
Db 467 LIDVIALNRYNGWYFDGDLLEAAKVHLR---QEFHAMNKRCPGKPIMITEYGADTVAGFH 523
QY 487 YDPQMFSEEEYQAEIVKTRILLKKDYIIGTHVWAFADFKTPQNVRRPILNHKGVFTRD 546
Db 524 DIDPVMFTEEEYQAEVNHVVFDEFENFVGEQAWNPFADFATSQGVMRVQGNKKGVFTRD 583
QY 547 RQPKLVAHVLRRLMSEV 563
Db 584 RKPCLAAHVFRERWTNI 600
```

## RESULT 10

```
US-10-364-649-18
; Sequence 18, Application US/10364649
; Publication No. US20030229921A1
; GENERAL INFORMATION:
; APPLICANT: Richard A. Jefferson and Jorge E. Mayer
; TITLE OF INVENTION: MICROBIAL B-GLUCURONIDASE GENES, GENE
; FILE REFERENCE: 190106.405C1
; CURRENT APPLICATION NUMBER: US/10/364,649
; CURRENT FILING DATE: 2003-02-12
; PRIOR APPLICATION NUMBER: 10/364,649
; PRIOR FILING DATE: 2003-02-12
; PRIOR APPLICATION NUMBER: US 09/270,957
; PRIOR FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 607
; TYPE: PRT
; ORGANISM: Bacillus sp.
; US-10-364-649-18
```

```
Query Match 30.8%; Score 925; DB 14; Length 607;
Best Local Similarity 35.7%; Pred. No. 3.3e-68;
Matches 220; Conservative 92; Mismatches 223; Indels 82; Gaps 15;

QY 4 PQRNKKRFFILLNGVWNLEV-----TSKDRPIAVPGSWNE--QYQDLCYEE 47
Db 9 PINTETRGVFDLNGVWNFKLDYKGLEEKWYESKLTDTISMAVPSSYNDIGVTKEIRNHI 68
QY 48 GPFTYKTTFFVPKXLSQKHIRLYFAAVNTDCEVFLNGEKVGENHIEYLPFEVDVTGKYS 107
Db 69 GYVWYEREFVPAVLKQDRIVLRFSGATHKAIIVYVNGELVVEHKGFLLPFEAEINNSLRD 128
QY 108 GENELRVVVENRLKVGFPSPKVPDSGTHTVGFFGS-----FPPANPDFFPYGGII 157
Db 129 GMRNVTVAVDNII-----DSTLTPVGLYSEHHEGLGKVIIRNKNPDFDFNYAGLH 178
QY 158 RPLVIEFTDHARILDIWVDTSESEPEKKLGKVKKIEVSEEA VGQEMTIKLG--EEKKI 215
Db 179 RPKVIYTPFTYVEDISVTDENGPR--TGTVTYTVDFQ---GKAETVKVSVVDEEGKV 231
```

```
QY 216 RTSNRFVEGEFILENARFWSLEDPLYPLKVELEKDEYTLDI-----GIRTISWDEKRLY 270
Db 232 VASTEGISGNVEIPNVILMEPLNTLYQIKVELVNDGLTIDVYEEBPGVRTVEVNDGKFL 291
QY 271 LNKGPVFLKGFKGHEEPVLGQGTFFPLMIKD FNLLKWINANSFRTSHYPYSEEWLAD 330
Db 292 INNKPFFYFKGFKGHEEDTPINGRGFNBSNVMDFNILKWINANSFRTAHYPYSEELMLAD 351
QY 331 RLGLIVIDEAPHVGITRHYHN-----PETQKIA-----EDNIRRMIDRHKN 371
Db 352 REGLVVIDETPAVG--HINFMATTGLGEGSERVSTWEKIRTFEHHQDVLRELVSBDKN 408
QY 372 HPSVIMSVANEPESNHPDAEGFFKALYETANEMD-RTRPVVMVSMMDAPDERTRDVALK 430
Db 409 HPSVIMWSIANEAATEEGAYEYFKPLVELTKELDPQKRPVTIVLFVWATPE--TDKVAE 466
QY 431 YFDIVCNRYYGWYIYQGRIEEGLQALEKDI EELYARHR---KPIFVTEFGADAIAGIH 486
Db 467 LIDVIALNRYNGWYFDGDLLEAAKVHLR---QEFHAMNKRCPGKPIMITEYGADTVAGFH 523
QY 487 YDPQMFSEEEYQAEIVKTRILLKKDYIIGTHVWAFADFKTPQNVRRPILNHKGVFTRD 546
Db 524 DIDPVMFTEEEYQAEVNHVVFDEFENFVGEQAWNPFADFATSQGVMRVQGNKKGVFTRD 583
QY 547 RQPKLVAHVLRRLMSEV 563
Db 584 RKPCLAAHVFRERWTNI 600
```

## RESULT 11

```
US-10-364-649-28
; Sequence 28, Application US/10364649
; Publication No. US20030229921A1
; GENERAL INFORMATION:
; APPLICANT: Richard A. Jefferson and Jorge E. Mayer
; TITLE OF INVENTION: MICROBIAL B-GLUCURONIDASE GENES, GENE
; FILE REFERENCE: 190106.405C1
; CURRENT APPLICATION NUMBER: US/10/364,649
; CURRENT FILING DATE: 2003-02-12
; PRIOR APPLICATION NUMBER: 10/364,649
; PRIOR FILING DATE: 2003-02-12
; PRIOR APPLICATION NUMBER: US 09/270,957
; PRIOR FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 28
; LENGTH: 615
; TYPE: PRT
; ORGANISM: Bacillus sp.
; US-10-364-649-28
```

```
Query Match 30.8%; Score 925; DB 14; Length 615;
Best Local Similarity 35.7%; Pred. No. 3.4e-68;
Matches 220; Conservative 92; Mismatches 223; Indels 82; Gaps 15;

QY 4 PQRNKKRFFILLNGVWNLEV-----TSKDRPIAVPGSWNE--QYQDLCYEE 47
Db 9 PINTETRGVFDLNGVWNFKLDYKGLEEKWYESKLTDTISMAVPSSYNDIGVTKEIRNHI 68
QY 48 GPFTYKTTFFVPKXLSQKHIRLYFAAVNTDCEVFLNGEKVGENHIEYLPFEVDVTGKYS 107
Db 69 GYVWYEREFVPAVLKQDRIVLRFSGATHKAIIVYVNGELVVEHKGFLLPFEAEINNSLRD 128
QY 108 GENELRVVVENRLKVGFPSPKVPDSGTHTVGFFGS-----FPPANPDFFPYGGII 157
Db 129 GMRNVTVAVDNII-----DSTLTPVGLYSEHHEGLGKVIIRNKNPDFDFNYAGLH 178
QY 158 RPLVIEFTDHARILDIWVDTSESEPEKKLGKVKKIEVSEEA VGQEMTIKLG--EEKKI 215
Db 179 RPKVIYTPFTYVEDISVTDENGPR--TGTVTYTVDFQ---GKAETVKVSVVDEEGKV 231
```





OY	278	LKGFGEHFEFVLGGTFYPLMIKDENLLKWINANSFRTSHYPYSEEWLDLADRLGLI VI	337
Dd	293	LKGFGKHEDFNVLGKAVNESIIKRDECKMKGIGANCFRSSHYPYAEEWYQYADKYGFL I I	352
OY	338	DEAPVVGITRYHYN-----PETOKIAEDNIRRMIDRHKNHPSVIAMS	379
Dd	353	DEVPAVGLNRSITNFLNTNSNQSHFPASKTVPBLKVYHEQEI KEMIDRDQRHPSVIAMS	412
OY	380	VANEPESNHPDAEGFEFKALYETANEND-RTRPVVMVSMMDAPDERTEVALKYFDIYCVN	438
Dd	413	LENPESTTOESYDFEKDIFAFARKUDPQNRPRYTGLVMGSgpK--VDKLHPLCDFVCIN	470
OY	439	RYYGWYIYQG-RIEGLQALEKDIIEELY-ARHRKPITEFGADAIAGIHYPDPOQMSZE	496
Dd	471	RYYGWYVAGSPEIVNAKKMLLEDLDGQNKLKNKPFVFTEFGADTLSSHRLPDEMWSQE	530
OY	497	YQAELEVETIRLLKKDYIIETHVMAFADFKTPQNVBRPI LNHKGVTFRDQPKLVAHYL	556
Dd	531	YQNEYYQMYPFDIFKKYRPFIGELVNMFADFKTSEGIMRVGVGNDKGIFTRDRBPKDIAFTL	590
OY	557	RRLMSEV 563	
Dd	591	KGRMOQL 597	

```

RESULT 14
US-10-421-175-4
; Sequence 4, Application US/10421175
; Publication No. US20030219414A1
; GENERAL INFORMATION:
; APPLICANT: Podsakoff, Gregory
; APPLICANT: Watson, Gordon
; APPLICANT: Couto, Linda B.
; APPLICANT: Yang, Bin
; TITLE OF INVENTION: RECOMBINANT ADENO-ASSOCIATED VIRUS VIRIONS FOR THE
; TITLE OF INVENTION: TREATMENT OF LYSOSOMAL DISORDERS
; FILE REFERENCE: 0800-0021
; CURRENT APPLICATION NUMBER: US/10/421,175
; CURRENT FILING DATE: 2003-04-22
; PRIOR APPLICATION NUMBER: US/09/715,858
; PRIOR FILING DATE: 2000-11-14
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 648
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-421-175-4

```

QY	252	EYTLDIGIRITSMDEKRLYLNGKPVFLKGFGEHIEFPVLQOGTFYPLMIKDFNLKWINA	311
		:: : :     : :     : :     : :     : :     : :	
Db	315	YTLPIGIRTVAVTKSKFLINGKRPFYQGVNKHEDSDIRGKGFDELLVKDFNLRLWIGA	374
QY	312	NSFRTSHYPYSEEWLIDLADRLGLVIDEAPHVGT-----RYHNPETOKIAEDN	361
Db	375	NSFRTSHYPYSEEVQLCDRYGIVIDECPGVGIVLPQSGFNESLRLHL----EYMEEL	429
QY	362	IRRMIDRHKHPSVIMSVANEBESNHPDAEGFEKALYETANEMDRTRPVMVS--NMDA	419
		:       : :     : :     : :     : :     : :     : :     : :	
Db	430	VR----RDKNHPAVVMMSVANEPSSALKPAAYYFXTLITHTKALDLTRPVTFSNAKYDA	485
QY	420	PDBERTRDVALKYFDIVCVNRYGYWYIYQGRIEBGLALEKDIIEELYARHRKPIFVTEGA	479
		:   : :     : :     : :     : :     : :     : :     : :	
Db	486	-----DLGAPYVDVICVNSYFSWYHDYGHLEVIQPOLNSQFENWYTKHQPIIOSEYGA	539
QY	480	DALAGIHYPDPQMFSESEYQALVEKTI RL L--LKKDYIIGTHVMAFADFKTPONVRRPIL	537
		:         : :     : :     : :     : :     : :	
Db	540	DALPGIHEDPPRMFSESEYQKAVLENYHSHVLDQKRREYVVGELIMNPFADFMTNQSPLRVIG	599
QY	538	NHKGVFTDRDQPKLVAVHLR-RLW 560	
		:     :     :     :     :     :     :     :     :     :	
Db	600	NKKGIFTROQPKTSAFILRERYW 623	

```

RESULT 15
US-09-118-276-12
: Sequence 12, Application US/09118276
: Patent No. US20010011381A1
: GENERAL INFORMATION:
: APPLICANT: BABYCHUK, ELENA;
: APPLICANT: KUSHNIR, SERGEI;
: APPLICANT: DE BLOCK, MARC;
: APPLICANT: INZE, DIRK
: TITLE OF INVENTION: METHODS AND MEANS TO MODULATE PROGRAMMED
: TITLE OF INVENTION: CELL DEATH IN EUKARYOTIC CELLS
: NUMBER OF SEQUENCES: 21
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: SIXBEY, FRIEDMAN, LEEDOM, & FERGUSON
: STREET: 8180 GREENSBORO DRIVE, SUITE 800
: CITY: MCLEAN,
: STATE: VIRGINIA
: COUNTRY: USA
: ZIP: 22102
: COMPUTER READABLE FORM:
: MEDIUM TYPE: 3-1/2" DISKETTE
: COMPUTER: IBM-COMPATIBLE
: OPERATING SYSTEM: MS-DOS
: SOFTWARE: ASCII
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/118,276
: FILING DATE: 17-JUL-1998
: PRIOR APPLICATION DATA: NONE
: ATTORNEY/AGENT INFORMATION:
: NAME: SCHULMAN, ROBERT M.; SALEN, KENNETH H.
: REGISTRATION NUMBER: 31,196; 43,077
: REFERENCE NUMBER: 6201-0003
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703) 790-9110
: TELEFAX: (703) 883-0370
: INFORMATION FOR SEQ ID NO: 12:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1010 RESIDUES
: TYPE: AMINO ACID
: STRANDEDNESS: SINGLE
: TOPOLOGY: LINEAR
:
US-09-118-276-12

```

Query Match	29.4%;	Score 882;	DB 9;	Length 1010;
Best Local Similarity	34.0%;	Pred. No. 2,7e-64;		
Matches 206;	Conservative 109;	Mismatches 229;	Indels 62;	Gaps 11;
0y	1	MWRPQNKKKR	FILLINGVNNLEV-----	TSKDRPIAVPGSNMEQYODLCY 45



```
Db 408 MVRPVEPTREIKKLDGLWAFSLDRENCGIDQRMWESALQESRAIAVPGSFNDQFADADI 467
QY 46 EE--GPFTYKTFYVPKXLSQKHRLYFAAVNTDCEVFLNGEKYGENHIEYLPEVDVTG 103
Db 468 RNYAGNVWYQREYFI PKGWAGQRLVLRFDVATHYKVMVNNQEVMEHQGYTPFEADVTP 527
QY 104 KYSGEN-ELRVVENRLKVGSPSKV--PDGTHTVGPFPGSPRANFDFFPYGGIIRP 159
Db 528 VYIAGKSVRITVCVNNELNMQTIPEGVITDENGKKQSYF-----HDFFNAGIHRS 580
QY 160 VLIETDARIIDIWDTSESEPEKKLGKVKVLEVSEAVGQEMTIKLGEEKKIRTSN 219
Db 581 VMLYTTPTNTWVDITVTHVAQ---DCNHASVDMQV---VANGDVSELRDADQVAVATG 634
QY 220 RVEGEFILENARFWSLEDPLYPL---KVELEKDEYTLDIGRTISWDEKRLYLNGKP 275
Db 635 QGTSGTLQVNVNPHLMQPGEGYLYELCVTAKSQTECDIYPLRVGIRSAVAVKGEQFLINHNP 694
QY 276 VFLKGFGKHHEFPVLGQTFYPLMIKDFNLKMINANSFRTSHYPYSEEWLDLADRLGITL 335
Db 695 FYFTGFGRHEDADLRGKGFDPNVLMVHDALMDWIGANSYRTSHYPYAEEMLDWADDEHGIV 754
QY 336 VIDEAPHVGI-----TRYHNPETOKIAEDNIRRMIDRHKNHPSVIM 377
Db 755 VIDETAAVGFNLSLGIGFEAGNKPEL YSEAVNGETQOAHLOAIKELIARDKNHPSVIM 814
QY 378 WSVANEPESNHPDAEGFFKALYETANEMDRTPVVMVSM--DAPDERTRDVALKYFDIV 435
Db 815 WSIANEPDTRPOGAREYFAPLAETRKLDPTRPITCVNMVFCDAHTDTISDL----FDVL 870
QY 436 CVNRYYGWYIYOGRIEGLQALEKDIEELYARHRKPIFVTEFGADAIAGIHYPQMFSE 495
Db 871 CLNRYYGWYVQSGDLETAEKVLEKELLAMQEKLHQPIITEYGVDTLAGLHSMYTDWSE 930
QY 496 EYQAELEVEKTRLLLLKDYIIGTHVWAFADFKTPQNVRRPILNHKGVFTDRQPKLVAAHV 555
Db 931 EYQCAWLDWYHRVFDRVSAVVGEOVMNFADFATSGILRVGKNKGI FTRDRKPKSAAPL 990
QY 556 LRRLWS 561
Db 991 LQKRWT 996
```

Search completed: January 24, 2005, 07:55:31  
Job time : 85 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 23, 2005, 22:33:20 ; Search time 854 Seconds  
(without alignments)  
10382.052 Million cell updates/sec

Title: US-09-936-759-14

Perfect score: 1689

Sequence: 1 atgctaagaccgcaacgaaa.....gaagactgtgagtgaggtc 1689

Scoring table: IDENTITY\_NUC

Searched: 4134886 seqs, 2624710521 residues

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_23Sep04:\*

1: geneseqn1980s:\*

2: geneseqn1990s:\*

3: geneseqn2000s:\*

4: geneseqn2001as:\*

5: geneseqn2001bs:\*

6: geneseqn2002as:\*

7: geneseqn2002bs:\*

8: geneseqn2003as:\*

9: geneseqn2003bs:\*

10: geneseqn2003cs:\*

11: geneseqn2003ds:\*

12: geneseqn2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1688	99.9	1689	3	AAA07937	Aaa07937 Thermotog
2	174	10.3	2128	4	ABL04147	Abi04147 Drosophil
3	174	10.3	2169	8	AAD50922	Aad50922 L. mexica
4	174	10.3	2169	12	ADF47502	Adf47502 Human bet
5	174	10.3	2169	12	ADJ58608	Adj58608 Fusion bet
6	172.4	10.2	1956	4	AAD06386	Aad06386 Human bet
7	172.4	10.2	2191	6	ABK83932	Abk83932 Human cdn
8	172.4	10.2	2191	10	ADD29602	Add29602 Human tum
9	172.4	10.2	2191	10	ADG89395	Adg89395 Cancer de
10	172.4	10.2	2191	12	ACF57506	Acf57506 Human bet
11	172.4	10.2	2191	12	ADJ62864	Adj62864 Human bet
12	172.4	10.2	2191	12	ADP10326	Adp10326 Reference
13	172.4	10.2	3525	4	AAF57574	Aaf57574 Mutant he
14	172.4	10.2	7328	4	AAF57573	Aaf57573 Heavy cha
15	170.8	10.1	3314	2	AAQ58896	Aaq58896 Humanised
16	166.8	9.9	2472	10	ADB58226	Adb58226 Toxicity-
17	166.8	9.9	2472	10	ADB52752	Adb52752 Primary r
18	159.6	9.4	1947	4	AAD06387	Aad06387 Murine be
19	145.8	8.6	2109	10	ADG76418	Adg76418 Human inc
20	142.2	8.4	23078	4	AAS59508	Aas59508 Propionib
21	142.2	8.4	23078	8	ACF64437	Acf64437 Proionib

22	141.4	8.4	1887	2	AAX23826	Aax23826 Bacillus
23	133	7.9	1820	3	AAA07958	Aaa07958 Staphyloc
24	132.6	7.9	1812	3	AAA07939	Aaa07939 Salmonell
25	130.4	7.7	1888	3	AAA07938	Aaa07938 Codon-opt
26	114	6.7	1814	10	ADC01686	Adc01686 Enterohae
27	114	6.7	3451	8	ABT16609	Abt16609 Artificia
28	114	6.7	3451	10	ACC44711	Acc44711 Plasmid p
29	114	6.7	14627	8	ABT16610	Abt16610 Artificia
30	114	6.7	14627	10	ACC44712	Acc44712 Plasmid p
31	110.8	6.6	1809	6	ABK87080	Abk87080 uidA DNA
32	110.8	6.6	1812	4	AAH22503	Aah22503 PCR gener
33	110.8	6.6	1812	6	AAD29543	Aad29543 uidA DNA
34	110.8	6.6	1812	6	ABL61371	Abi61371 GUS DNA #
35	110.8	6.6	1812	8	ABT16606	Abt16606 Artificia
36	110.8	6.6	1812	10	ACC44709	Acc44709 E. coli b
37	110.8	6.6	2001	4	AAD19830	Aad19830 Beta-gluc
38	110.8	6.6	2001	6	ABK15667	Abk15667 Glutathio
39	110.8	6.6	2186	3	AAZ38598	Aaz38598 Dysfuncti
40	110.8	6.6	2457	1	AAH81451	Aan81451 Sequence
41	110.8	6.6	2673	1	AAH81449	Aan81449 Sequence
42	110.8	6.6	2725	4	AAD19835	Aad19835 Promoter-
43	110.8	6.6	2730	4	AAD19833	Aad19833 Promoter-
44	110.8	6.6	3035	2	AAV37748	Aav37748 PAT1 gene
45	110.8	6.6	3169	2	AAT38397	Aat38397 PTIC 99 p

ALIGNMENTS

RESULT 1

AAA07937

ID AAA07937 standard; DNA; 1689 BP.

XX

AC AAA07937;

XX

DT 26-JAN-2001 (first entry)

XX

DE Thermotoga maritima beta-glucuronidase gene.

XX

KW Microbial; beta-glucuronidase; GUS; Enterobacter; Salmonella;

KW Pseudomonas; Staphylococcus; Thermotoga; transgenic plant; bioindicator;

KW transgenic insect; marker; glucuronide detoxification; ds.

XX

OS Thermotoga maritima.

XX

PN WO200055333-A1.

XX

PD 21-SEP-2000.

XX

PF 16-MAR-2000; 2000WO-US007107.

XX

PR 17-MAR-1999; 99US-00270957.

XX

PA (CAMB-) CAMBIA BIOSYSTEMS LLC.

XX

PI Jefferson RA, Mayer JE;

XX

DR WPI: 2000-647075/62.

DR P-PSDB; AAB28406.

XX

PT Novel microbial beta-glucuronidase genes and gene products used as

PT reporter/effector molecule, as diagnostic tool, in positive selection, to

PT target molecules to specific cells and to detect and track linked genes.

XX

PS Claim 2; Fig 4; 116pp; English.

XX

CC The present sequence encodes a microbial beta-glucuronidase (GUS)

CC protein. GUS genes were obtained from six different genera:

CC Enterobacter/Salmonella, Pseudomonas, Salmonella, Staphylococcus and

CC Thermotoga. Microbial GUS can be used as a reporter/effector molecule for

CC transgenic constructions and in in vitro diagnostic applications. It may

CC also be used to generate sentinel plants that serve as bioindicators of

CC environmental status. It may be used to generate transgenic insects for

tracking insect populations or to facilitate the development of a bioassay for compounds that affect molecules critical for insect development (e.g. juvenile hormone). Secreted GUS may also serve as a marker for beneficial fungi destined for release into the environment. In animal systems, secreted GUS may be used to achieve extracellular detoxification of glucuronides (e.g. toxin glucuronide) and to examine conjugation patterns of glucuronides. Microbial GUS may also be used in traditional medical diagnostic assays, for drug testing, pharmacokinetic studies, bioavailability studies, diagnosis of diseases and syndromes, following progression of disease or its response to therapy. Microbial GUS has increased thermal stability, high turnover number and enzymatic activity. It is highly specific for the substrate and water soluble, and the substrates are stable

SQ Sequence 1689 BP; 526 A; 374 C; 430 G; 358 T; 0 U; 1 Other;

Query Match	99.9%;	Score 1688;	DB 3;	Length 1689;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 1689;	Conservative	0;	Mismatches	0;
			Indels	0;
			Gaps	0;

QY	I	ATGTAAGACCCGCAACGAACACAGAAAGATTATTTATCTTATCTTGAATGAGTTTGAAT	60
Db	1	ATGTAAGACCCGCAACGAACACAGAAAGATTATTTATCTTATCTTGAATGAGTTTGAAT	60
QY	61	CTTGAAGTAACCAAGACAAGACAGCAACCAATCGCCGTTCTGGAAGCTGGAATGAGAGTAC	120
Db	61	CTTGAAGTAACCAAGACAAGACAGCAACCAATCGCCGTTCTGGAAGCTGGAATGAGAGTAC	120
QY	121	CAGATCTGTGCTACGAAGAAGGACCTTCACTCAAAAAACCACTTCTACGTTCCGAAG	180
Db	121	CAGATCTGTGCTACGAAGAAGGACCTTCACTCAAAAAACCACTTCTACGTTCCGAAG	180
QY	181	NAACTTTACAACAAAAACATCAGACTTTACTTTGTGCGGTGAACACGACTGCGAGGTC	240
Db	181	NAACTTTACAACAAAAACATCAGACTTTACTTTGTGCGGTGAACACGACTGCGAGGTC	240
QY	241	TTCTCTCAACGAGAGAAAAGTGGGAGAGATCACATTGAATACCTTCCCTTGAAGTAGAT	300
Db	241	TTCTCTCAACGAGAGAAAAGTGGGAGAGATCACATTGAATACCTTCCCTTGAAGTAGAT	300
QY	301	GTGACGGGGAAAAGTGAATCCGGAGAGAACGAACCTCAGGGGTGTTGTGAGAACAGATTG	360
Db	301	GTGACGGGGAAAAGTGAATCCGGAGAGAACGAACCTCAGGGGTGTTGTGAGAACAGATTG	360
QY	361	AAAGTGGGAGGATTTCCCTCGAAGGTTCCAGACAGCGGCACTCACACCGTGGATTTTTT	420
Db	361	AAAGTGGGAGGATTTCCCTCGAAGGTTCCAGACAGCGGCACTCACACCGTGGATTTTTT	420
QY	421	GGAAGTTTTCCACCTGCAAACTTGCATTTCTCCCTACGGTGAATCATAGGCGCTGTT	480
Db	421	GGAAGTTTTCCACCTGCAAACTTGCATTTCTCCCTACGGTGAATCATAGGCGCTGTT	480
QY	481	CTGATAGAGTTCAAGACCAACGCGAGGATCTGCACATCTGGGTGACACGAGTGAAGTCT	540
Db	481	CTGATAGAGTTCAAGACCAACGCGAGGATCTGCACATCTGGGTGACACGAGTGAAGTCT	540
QY	541	GAAACGGAGAGAAGAACTTGAAAAAGTGAAAGTGAAGATCTCAGAAAGACGGTG	600
Db	541	GAAACGGAGAGAAGAACTTGAAAAAGTGAAAGTGAAGATCTCAGAAAGACGGTG	600
QY	601	GGAACGAGATGACGATCAAACTTGAGAGAGAAAGAAAAAGATTGAACATCCAACAGA	660
Db	601	GGAACGAGATGACGATCAAACTTGAGAGAGAAAGAAAAAGATTGAACATCCAACAGA	660
QY	661	TTTCGTGAAAGGGAGTTTCCTCTGAAAAACGACAGTTCTGAGCCTCGAAGATCCATAT	720
Db	661	TTTCGTGAAAGGGAGTTTCCTCTGAAAAACGACAGTTCTGAGCCTCGAAGATCCATAT	720
QY	721	CTTTATCTCTCAAGGTGAACTTGAAAAAGACGAGTACACTCTGACATCGGAATCAGA	780
Db	721	CTTTATCTCTCAAGGTGAACTTGAAAAAGACGAGTACACTCTGACATCGGAATCAGA	780
QY	781	ACGATCAGCTGGGACGAGAAAGAGGCTTATCTGAAACGGGAAAACCTGTCTTTTGAAGGCG	840

Db	781	ACGATCAGCTGGGACGAGAAAGGCTCTATCTGAA	CGGGAACCTGCTTTTGAAGGC	840	
QY	841	TTTGGAAAGCAGAGGAATTC	CCCGTTCTGGGGCAGGGCACTTTATCCATTGATGATA	900	
Db	841	TTTGGAAAGCAGAGGAATTC	CCCGTTCTGGGGCAGGGCACTTTATCCATTGATGATA	900	
QY	901	AAAGACTTCAACTTCTGAAAGTGATCAACGCCGAATTTCTTTCAGGACCTCTCACTATCCT		960	
Db	901	AAAGACTTCAACTTCTGAAAGTGATCAACGCCGAATTTCTTTCAGGACCTCTCACTATCCT		960	
QY	961	TACAGTGAAGAGTGGCTGGATCTTGCCGACAGACTCGGAATCCTTGTA	TAGACGAAGCC	1020	
Db	961	TACAGTGAAGAGTGGCTGGATCTTGCCGACAGACTCGGAATCCTTGTA	TAGACGAAGCC	1020	
QY	1021	CCGCACGTTGGTATCACAAGGTACCACTACATCCCGAGACTCAGAAAGATAGCAGAAGAC		1080	
Db	1021	CCGCACGTTGGTATCACAAGGTACCACTACATCCCGAGACTCAGAAAGATAGCAGAAGAC		1080	
QY	1081	AACATAAGAGAATGATCGACAGACACAGAACCATCCCACTGTGATCATGTGAGTGTG		1140	
Db	1081	AACATAAGAGAATGATCGACAGACACAGAACCATCCCACTGTGATCATGTGAGTGTG		1140	
QY	1141	GCGAACGAACCAAGAGTCCAA	CCATCCAGAGCGGAGGTTCTTCAAGCCCTTATGAG	1200	
Db	1141	GCGAACGAACCAAGAGTCCAA	CCATCCAGAGCGGAGGTTCTTCAAGCCCTTATGAG	1200	
QY	1201	ACTGCCAATGAATGATCGAACACGCCCCGTTGT	CATGGTGAAGCATGATGACGCACCA	1260	
Db	1201	ACTGCCAATGAATGATCGAACACGCCCCGTTGT	CATGGTGAAGCATGATGACGCACCA	1260	
QY	1261	GACGAGAGAACAGAGACGTGGCGTGAAGTACTTGCACATCGTCTGTGTGAACAGGTAC		1320	
Db	1261	GACGAGAGAACAGAGACGTGGCGTGAAGTACTTGCACATCGTCTGTGTGAACAGGTAC		1320	
QY	1321	TACGGCTGTACATCTATCAGGGAAAGATAGAAAGAGACTTCAAGCTCTGGA	AAAAAGAC	1380	
Db	1321	TACGGCTGTACATCTATCAGGGAAAGATAGAAAGAGACTTCAAGCTCTGGA	AAAAAGAC	1380	
QY	1381	ATAGAAGAGCTCTATGCAAGGCA	CAGAAAGCCCATCTTTGTCA	CAGAATTGCTGCGGAC	1440
Db	1381	ATAGAAGAGCTCTATGCAAGGCA	CAGAAAGCCCATCTTTGTCA	CAGAATTGCTGCGGAC	1440
QY	1441	GCGATAGCTGGCATCCACTACGATCCACCTCAAA	TGTTCTCCGAAGAGTACCAAGCAGAG	1500	
Db	1441	GCGATAGCTGGCATCCACTACGATCCACCTCAAA	TGTTCTCCGAAGAGTACCAAGCAGAG	1500	
QY	1501	CTCGTTGAAAAAGACGATCAGGCTC	TTTTGAAAAAAGACTACATCATCGGAACACACGTG	1560	
Db	1501	CTCGTTGAAAAAGACGATCAGGCTC	TTTTGAAAAAAGACTACATCATCGGAACACACGTG	1560	
QY	1561	TGGGCTTTTGACAGATTTTAA	GACTCTCAGAAATGTGAGAAAGACCCATTCTCAA	CCACAAG	1620
Db	1561	TGGGCTTTTGACAGATTTTAA	GACTCTCAGAAATGTGAGAAAGACCCATTCTCAA	CCACAAG	1620
QY	1621	GGTGTTTTTCACAAGAGACAGACAACCCAA	CTCGTTGCTCATGTACTGAGAAAGACTGTGG	1680	
Db	1621	GGTGTTTTTCACAAGAGACAGACAACCCAA	CTCGTTGCTCATGTACTGAGAAAGACTGTGG	1680	
QY	1681	AGTGAAGTT	1689		
Db	1681	AGTGAAGTT	1689		

RESULT 2	
ABL04147	
ID	ABL04147 standard; cDNA, 2128 BP.
XX	
AC	ABL04147;
XX	
DT	26-MAR-2002 (first entry)
XX	
DE	Drosophila melanogaster expressed polynucleotide SEQ ID NO 6923

XX Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical; gene; ss.  
XX  
OS Drosophila melanogaster.  
XX  
PN WO200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 23-MAR-2001; 2001WO-US009231.  
XX  
PR 23-MAR-2000; 2000US-0191637P.  
PR 11-JUL-2000; 2000US-00614150.  
XX  
PA (PEKE ) PE CORP NY.  
XX  
PI Venter JC, Adams M, Li PWD, Myers EW;  
XX  
DR WPI; 2001-656860/75.  
DR P-PSDB; ABB60044.  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signaling and cell-cell  
PT interactions.  
XX  
PS Claim 1; SEQ ID NO 6923; 21pp + Sequence Listing; English.  
XX  
CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA  
CC sequences (AB101840-AB16175) and the encoded proteins (ABB57737-  
CC ABB72072). The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 2128 BP; 546 A; 511 C; 540 G; 531 T; 0 U; 0 Other;

Query Match 10.3%; Score 174; DB 4; Length 2128;  
Best Local Similarity 51.6%; Pred. No. 2.5e-43;  
Matches 492; Conservative 0; Mismatches 435; Indels 27; Gaps 3;

QY 734 AGGTGGAACCTTGAAAAAGACGATACACTCTGGACATCGGAATCAGAACGATCAGCTGGG 793  
DB 1104 AGTTGGAATCTCTGCAGAGATACCTACCGCTCCAGTGGGCAATACGTAAAGTTGGG 1163  
QY 794 ACGAGAAAGAGGCTCTATCTGAAAGGGAACCTGTCTTTTGAAGGGCTTTGGAAGCAGC 853  
DB 1164 ACAACGACAGCCTGTCTGTGAATGGAACCAACCTCTATCTGCGGGATTGACGCGCAGC 1223  
QY 854 AGGAATTCCTCCGTTCTGGGGCAGGGCACCTTTATTCATTGATGATTAAGACTTCAACC 913  
DB 1224 AGGACTCCGATATCCGCGGGAGGGATTGGATTAATGCGTTCTTGCTAGAGATTTAAAC 1283  
QY 914 TTCTGAAGTGATCAACGCGGAATCTTTACGACCTCTCACTATCCTTACAGTGAAGAGT 973  
DB 1284 TGCTGAAGTGAGCTGAGCCAAATGCATATCGACCTCTCACTATCCTTATTCGAAAGAGT 1343  
QY 974 GGCTGATCTTGCCGACAGACTCGGAATCCTTGATAGACGAAGCCCCGACGTTGTA 1033  
DB 1344 CAATGCAATTTGCCGATCAGCATGTATATTGATTAATTGACGAATGCC-----TGCTG 1397  
QY 1034 TCACAAGGTACACTACATCCGAGACTCAGAGATAGCAGAAAGACAACATTAAGAGAA 1093  
DB 1398 TCAATATAGATATCTTCGAGCCGACAGTACTGAGAGATCACAATGTCTCGTGAACAAC 1457  
QY 1094 TGATGCACAGACAAGAACCATCCAGTGTGATCATGTGAGTGTGGCAAGAACAG 1153  
DB 1458 TGATCCACCGGAGACAGAAACCATCCAAAGTGTGTGATGTGCGTAGCCAAAGAGCCGA 1517

QY 1154 AGTCCAACCATCCAGACGCGGAGGGTTCTTCAAGCCCTTTATGAGACTGCCAATGAAA 1213  
DB 1518 GATCGAACAGACAGGAGCCCTTAATACCTTGAATTCCTGGTAACATGTAAGAGAAA 1577  
QY 1214 TGGATCGAACAGCCCCGTTGTCTATGTGATGATGATGACGACACGACGAGAACMA 1273  
DB 1578 TAGCTCACGAGACAGCTTAAACCGCGCTAATAATGCCAACTC-----TTCCA 1625  
QY 1274 GAGACGTGGCGCTGAAGTACTTGCACATCGTCTGTGTGAACAGTACTACGCTGTACA 1333  
DB 1626 GTTGCCATTGGCGCAGTTTCTGACATCGTGGGTTTCAATCGCTATTAATCTTGTATC 1685  
QY 1334 TCTATCAGGGAAGATAGAAAGAGACTTCAAGCTCTGAAAAAGACATAGAGAGCTCT 1393  
DB 1686 AGAACTCGGCGCTACTGACATGATACCTTAACTGTTGACGATAGAGCCAGAGTTGGC 1745  
QY 1394 ATGCAAGGCACAGAAAGCCCATCTTTGTACAGAAATTCGGTGGACGCGATAGCTGCA 1453  
DB 1746 GGGATAGTTGCGAAAGCCTGTATCAATTTGAGTACGCGCGGACACTAATGAGGGCA 1805  
QY 1454 TCCACTACGATCCACCTCAATGTCTCCGAGAGTACCAAGCAGAGCTCGT----- 1505  
DB 1806 TGCACACTCTCCCGCCTTTATTTGGTGGAGGAATACCAAGTTGAGCTCTTCTCGCC 1865  
QY 1506 -TGAAAGACGATCAGGCTCTTTGAAAAAAGACTACATCATCGGAACACACGTGTGG 1564  
DB 1866 ATTTCAAGGCTTTGACAGATTAACGCGAAGAGATGTTAATTGAGAGTTGTTTGA 1925  
QY 1565 CTTTGCAGATTTTAAGACTCTCAGAAATGAGAAAGACCATTCTCAACCAAGAGGTG 1624  
DB 1926 ACTTCGCCGATTTTCGACGCGCAGACATTAACCGCGTGGCGGCAACAAAAGGAG 1985  
QY 1625 TTTTCAAGAGACAGACAAACCAACTCGTTGCTCATGTACTGAGAAGACTGT 1678  
DB 1986 TCTTACAAGGAACCGAACCAACCAAGAGTGGCTCACATTTAGGCGGCGGT 2039

RESULT 3  
AAD50922  
ID AAD50922 standard; DNA; 2169 BP.  
XX  
AC AAD50922;  
XX  
DT 02-APR-2003 (first entry)  
XX  
DE L. mexicana SAP-human beta-GUS-IGF-II chimeric DNA.  
XX  
KW Lysosome; metabolic disease; lysosomal storage disease; gene therapy;  
KW Gaucher's disease; Pompe's disease; Hurler's syndrome; neuroprotective;  
KW Niemann-Pick's disease; Schindler's disease; mucopolidosis; cystinosis;  
KW Batten's disease; prosaposin; infantile neuronal ceroid lipofiscinosis;  
KW fucosidosis; mannosidosis; antilipæmic; insulin-like growth factor 2;  
KW IGF-II; secreted acid phosphatase; SAP; beta-glucuronidase; GUS; human;  
KW chimeric; gene; ds.  
XX  
OS Homo sapiens.  
OS Leishmania mexicana.  
OS Unidentified.  
OS Chimeric.  
XX  
FH Key Location/Qualifiers  
FT CDS 1..2169  
FT /\*tag= a  
FT /product= "L. mexicana SAP-human beta-GUS-IGF-II chimeric  
FT protein"  
FT 1..96  
FT /\*tag= b  
FT /note= "Leishmania mexicana SAP peptide"  
FT mat\_peptide 97..1806  
FT /\*tag= c  
FT /product= "Human mature beta-GUS protein"  
FT 1807..1815  
FT misc\_feature  
FT /\*tag= d



FT /note= "Linker"  
FT m1sc\_feature 1816..2169  
FT /\*tag= e  
FT /note= "IGF-II sequence"  
XX  
PN MO200287510-A2.  
XX  
PD 07-NOV-2002.  
XX  
PF 30-APR-2002; 2002WO-US013835.  
XX  
PR 30-APR-2001; 2001US-0287531P.  
PR 10-JUL-2001; 2001US-0304609P.  
PR 15-OCT-2001; 2001US-0329461P.  
PR 23-JAN-2002; 2002US-0351276P.  
XX  
PA (SYMB-) SYMBIONTICS INC.  
XX  
PI Lebowitz JH, Beverley SM;  
XX  
DR WPI; 2003-111838/10.  
DR P-PSDB; AAE33322.  
XX  
XX  
PT New targeted therapeutic that is active in a mammalian lysosome binds an  
PT extracellular domain of human cation-independent mannose-6-phosphate  
PT receptor, useful for treating metabolic diseases such as lysosomal  
PT storage disease.  
XX  
XX  
PS Example 1, Fig 3; 68pp; English.  
XX  
XX The invention relates to targeted therapeutic comprising a therapeutic  
CC agent that is active in a mammalian lysosome, and a means for binding an  
CC extracellular domain of human cation-independent mannose-6-phosphate  
CC receptor in a mannose-6-phosphate independent manner. The invention is  
CC useful for treating metabolic diseases such as lysosomal storage disease,  
CC e.g. Gaucher's disease, Pompe's disease, Hurler's syndrome, Niemann-  
CC Pick's disease, mannosidosis, fucosidosis, Schindler's disease,  
CC mucopolidosis, cystinosis, Batten's disease, prosaposin, or infantile  
CC neuronal ceroid lipofuscinosis. The invention is also useful in gene  
CC therapy. The present sequence is a DNA encoding Leishmania mexicana  
CC secreted acid phosphatase (SAP) signal peptide, human mature beta-CUS  
CC (glucuronidase), linker and IGF-II chimeric protein. This sequence is  
CC used in the exemplification of the invention  
XX  
SQ Sequence 2169 BP, 468 A; 609 C; 637 G; 455 T; 0 U; 0 Other;  
  
Query Match 10.3%; Score 174; DB 8; Length 2169;  
Best Local Similarity 52.2%; Pred. No. 2.5e-43;  
Matches 493; Conservative 0; Mismatches 430; Indels 21; Gaps 4;  
  
QY 737 TGGAACTTAAAAAGACGAGTACACTCTGGACATCGGAATCAGACGATCAGCTGGAGC 796  
DB 941 TGGGGCCCTGTGTCTGACTTCTACACACTCCCTGTGGGATCCGACATGCTGTACCA 1000  
QY 797 AGAAGAGGCTCTATCTGAACGGGAAACCTGTCTTTTGAAGGGCTTTGAAAGACAGAGG 856  
DB 1001 AGAGCCAGTTCCTCATCAATGGGAAACCTTTCTATTCCACGGTGTCAACAAGCATGAGG 1060  
QY 857 AATTCCCCGTTCTGGGGGAGGGACCTTTTATCCATTGATGATAAAGACTTCAACCTTC 916  
DB 1061 ATGCGGACATCCGAGGGAAGGGCTTCACTGGCCGCTGCTGTGAAGACTTCAACCTGC 1120  
QY 917 TGAAGTGAATCAACGCGAATTCTTTCAGAGCCTCTCACTATCTTTCAGAGTGAAGAGTGGC 976  
DB 1121 TTCGCTGGCTTGTGTGCCAAGCCTTTCCTTCCGTAACAGCCACTAACCCCTTATGCAAGAGGTGA 1180  
QY 977 TGGATCTTGGCCGACAGACTCGGAATCCTTGTGATAGAGAAAGCCCGCAGCTTGTATCA 1036  
DB 1181 TGCAGATGTGTGACCGGCTATGGATTGTGTCATGATGATGTATCCCGGCGTGGTCTGG 1240  
QY 1037 CAAGGTACCA---CTACAATCCCGAGACTCAGAAAGATAGCAGAAACAACATTAAGAGAA 1093  
DB 1241 CGCTGCCGCAAGTTCTTCAACAACGTTTCTCTGCATCAACCATGATGATGATGAAGAAG 1300

QY 1094 TGATGCACAGACACAGAAGCAACCATCCAGTGTGATCATGTGAGTGTGGCGAAGCAACCA 1153  
DB 1301 TGTGCGTAGGAGCAAGAACCAACCCCGCGGTCTGATGTGTGTGGCCCAACGAGCCTG 1360  
QY 1154 AGTCCAAACCATCCAGACCGCGGAGGTTTCTTCAAGCCCTTTATGAGACTGCGCAATGAAA 1213  
DB 1361 CGTCCACCTAGAAATCTGCTGGCTACTACTTGAAGATGATGATGCTCACAACCAATCCT 1420  
QY 1214 TGGATCGAACAACGCCCGTGTTCATGTGTGAGCATGATGAGCCAGACGAGAGAACAA 1273  
DB 1421 TGAACCCCTCCCGGCTGTGACCTTTGTAGCAACTTAATATGCAAG----- 1468  
QY 1274 GAGACGTGGCGCTGAAGTACTTCGACATCTGTGTGAAACAGTACTACGGCTGTGACA 1333  
DB 1469 CAGACAAAGGGGCTCCGTATGTGATGTGATCTGTTTGAACAGCTACTACTCTGGTATC 1528  
QY 1334 TCTATCAGGGAAGATAGAAAGAACTTCAAGCTCTGMAAAAACATAGAAAGCTCT 1393  
DB 1529 ACGACTACGGGCACTGAGTTGATTCAGCTGCACTGGCCACCAAGTTGAGAACTGGT 1588  
QY 1394 ATGCAAGGCACAGAAAGCCCATCTTTGTACACAAATTCGTGGGACGCGATAGCTGCA 1453  
DB 1589 ATAAAGAGTATCAGAAAGCCCATTAATTCAGAGCGAGTATGAGAGAAACGATTCAGAGGT 1648  
QY 1454 TCCACTACGATCCACTCAATGTCTCTCCGAAAGATACCAAGCAGAGCTCGTTGAAAAG- 1512  
DB 1649 TTGACCAAGATCCACTCTGATGTTCACTGAAAGATACCAAGAAAGTCTGTAGACAGT 1708  
QY 1513 ACGATCAGGCTCCTTTGAAAAA----AGACTACATATCGGAACACACGCTGGGCT 1567  
DB 1709 ACCATCTGGGTCTGGATCAAAAAACGAGAAATATGTGTTGAGAGCTCATTTGGAATT 1768  
QY 1568 TTGCAAGATTTTAAAGCTCCTCAGAAATGTGAGAAAGCCCATTTCAACCAAGAGGTGTTT 1627  
DB 1769 TTGCCGATTTTCAATGACTGAACAGTCAACCGCAGAGAGTGTGGGAAATAAAAAGGGATCT 1828  
QY 1628 TCACAAGAGACAGCAACCCAACTGTTGCTCATGTACTGAGA 1671  
DB 1829 TCACTCGGCAGAGACAACCAAAAGTGACGCTTCTTTTGGCA 1872  
  
RESULT 4  
ADF47502  
ID ADF47502 standard; cDNA; 2169 BP.  
XX  
AC ADF47502;  
XX  
DT 26-FEB-2004 (first entry)  
XX  
DE Human beta-glucuronidase and IGF-II fusion protein cDNA SEQ ID NO:5.  
XX  
XX underglycosylated targeted therapeutic; human; lysosome;  
KW lysosomal targeting domain;  
KW cation-independent mannose-6-phosphate receptor;  
KW underglycosylated therapeutic fusion protein; nephrotropic;  
KW enzyme replacement therapy; cell therapy; gene therapy;  
KW lysosomal storage disease; metabolic disease; enzyme defect;  
KW Gaucher disease; Tay-Sachs disease; Sandhoff disease; Fabry disease;  
KW Hunter syndrome; Sly syndrome; Schindler disease;  
KW infantile sialic acid storage disease; Batten disease;  
KW infantile neuronal ceroid lipofuscinosis; Ehlers-Danlos syndrome type VI;  
KW congenital glycosylation disorder; fusion protein; gene; ss.  
XX  
OS Synthetic.  
OS Homo sapiens.  
XX  
PN WO2003102583-A1.  
XX  
PD 11-DEC-2003.  
XX  
PF 29-MAY-2003; 2003WO-US017211.

[illegible]

Db	1121	TTGCGTGGCTTGGTGGCCAAACGCTTTCGGTAACGACCACTAACCCCTATGACAGAGAAAGTGA	1180		
QY	977	TGGATCTTGGCCGACAGACTCGGAATCCTTGTGATAGACGAAGCCCGCAGCTGGTATCA	1036		
Db	1181	TGCAGATGTGTGACCGCTATGGGAATGTGTGTCATGATGATGTCTCCCGGTGGTCTGG	1240		
QY	1037	CAAGGTACCA---CTACAATCCGAGACTCAGAAAGATAGCAGAAACAATAGAAACA	1093		
Db	1241	CGCTCCGCGACGTTCTTCAACAAAGTTTCTCTGCATCACCAATGCAGGTGATGAAAG	1300		
QY	1094	TGATCGACAGACAAAGAACATCCAGTGTGATCATGTGGAGTGTGGCGAAAGAACAG	1153		
Db	1301	TGTGTGCGTAGGGACAAAGAACACCCCGCGGTGTGATGTGTGTCTGTGGCCAAAGAGCTTG	1360		
QY	1154	AGTCCAAACCATCCAGACGCGGAGGGTTCTTCAAAAGCCCTTATGAGACTGCCAATGAAA	1213		
Db	1361	CGTCCACCTAGAAATCTGTCTGGCTACTACTTGAAGATGTGATGCTTCACACCAATCTT	1420		
QY	1214	TGGATCGAAACAGCCCCGTGTGTATGTGTGATGATGAGCGCACCGACGAGAGAAACA	1273		
Db	1421	TGGACCCCTCCCGGCTGTGACCTTTGTGTAGCAACTTAATATATGACAG-----	1468		
QY	1274	GAGACGTGGCGCTGAAGTACTTGCACATCGTCTGTGTGACAGGTACTACGGCTGTACA	1333		
Db	1469	CAGACAAAGGGGCTCCGTATGTGATGTGATGTGATCTGTTGAACAGCTACTACTCTTGGTATC	1528		
QY	1334	TCTATCAGGGAAGGATAGAAAGAGACTTCAAGCTCTGGAAGAAAGACATAGAAAGACTCT	1393		
Db	1529	ACGACTACGGGCACTGTGAGTTGATTACGTGCAGCTGGCCACCAGTTTGAAGAACTGGT	1588		
QY	1394	ATGCAAGGCACAGAAAGCCCATCTTTGTGACAGAAATTCGGTGCAGCGGATAGCTGGCA	1453		
Db	1589	ATAAGAAATATCAGAAAGCCCATTTATTACAGAGCGAGTATGAGCAGAAACGATTGCAGGGT	1648		
QY	1454	TCCACTACGATCCACCTCAAATGTTCTCCGAAGAGTATCCAGACGAGCTGTTGAAAAG-	1512		
Db	1649	TTCACCAAGATCCACCTCTGATGTTCATCTGAAAGATACAGAAAGTCTGTAGAGCAGT	1708		
QY	1513	ACGATCAGGCTCTTTGAAAAA-----AGACTACATCATCGGAACACAGCTGTGGGCTT	1567		
Db	1709	ACCATCTGGCTGTGATCAAAAAACGAGAAATATGTGTGGAGAGCTCATTTGGAATT	1768		
QY	1568	TTGCAGATTTTAAAGACTCCTCAGAAATGTGAGAAAGACCATTTCTCAACCAAGGGGTGTTT	1627		
Db	1769	TTGCCGATTTCATGACTGAAACAGTACCGACGAGAGTGTGGGAATAAAAAGGGGATCT	1828		
QY	1628	TCACAAAGAGACAGACAACCCAAACTCGTGTCTCATGTACTGAGA	1671		
Db	1829	TCACTCGGACAGACAAACCAAAAAAGTGACCGTTCCTTTTGGCA	1872		
RESULT 5					
ID	ADJ58608	standard; DNA; 2169 BP.			
XX	ADJ58608;				
AC					
XX					
DT	06-MAY-2004	(first entry)			
XX					
DE	Fusion DNA used in the exemplification of the invention.				
XX					
KM	Targetted therapeutic; lysosome storage disease; gene therapy; human;				
KW	IGF-II; insulin-like growth factor 2; beta-glucuronidase; GUS; chimeric;				
KM	gene; ds.				
XX					
OS	Homo sapiens.				
OS	Chimeric.				
OS	Unidentified.				
XX					
FH	Key	Location/Qualifiers			
FT	CDS	1..2169			
FT		/*tag= b			
FT		/product= "Fusion protein"			

```
FT      misc_feature      1..69
FT      /tag= a
FT      /note= "Signal peptide DNA"
FT      misc_feature      70..1956
FT      /tag= c
FT      /note= "Human mature P-glucoronidase DNA"
FT      misc_feature      1957..1965
FT      /tag= d
FT      /note= "Encodes an amino acid bridge"
FT      misc_feature      1966..2166
FT      /tag= e
FT      /note= "Human IGF-II DNA"
PN      US2004006008-A1.
XX      08-JAN-2004.
XX      16-OCT-2002; 2002US-00272483.
XX      30-APR-2001; 2001US-0287531P.
XX      10-JUL-2001; 2001US-0304609P.
XX      15-OCT-2001; 2001US-0329461P.
XX      23-JAN-2002; 2002US-0351276P.
XX      30-APR-2002; 2002US-00136841.
XX      29-MAY-2002; 2002US-0384452P.
XX      05-JUN-2002; 2002US-0386019P.
XX      06-SEP-2002; 2002US-0408816P.
PA      (SYMB-) SYMBIONICS INC.
XX      Lebowitz JH, Beverley SM;
XX      WPI; 2004-081736/08.
XX      P-PSDB; ADJ58609.
DR      New targeted therapeutic comprising a therapeutic agent that is active in
XX      mammalian lysosome, and means for binding a domain of human cation-
PT      independent mannose-6-phosphate receptor, useful for treating lysosome
PT      storage disease.
XX      Example 1; SEQ ID NO 5; 46pp; English.
PS      The invention relates to a targeted therapeutic which comprises a
XX      therapeutic agent that is active in a mammalian lysosome and means for
XX      binding an extracellular domain of human cation-independent mannose-6-
XX      phosphate receptor in a mannose-6 phosphate-independent manner. The
XX      targeted therapeutic, therapeutic fusion protein and methods of the
XX      invention are useful for treating lysosome storage diseases. The
XX      invention is also useful in gene therapy. The present sequence is a
XX      fusion DNA used in the exemplification of the invention. This DNA encodes
XX      a fusion protein which comprises a signal peptide sequence, a human
XX      mature beta-glucuronidase (GUS) sequence, a bridge of three amino acids
XX      and a human IGF-II sequence.
SQ      Sequence 2169 BP; 468 A; 609 C; 637 G; 455 T; 0 U; 0 Other;
XX      Query Match      10.3%; Score 174; DB 12; Length 2169;
XX      Best Local Similarity 52.2%; Pred. No. 2.5e-43;
XX      Matches 493; Conservative 0; Mismatches 430; Indels 21; Gaps 4;
QY      737 TGAACCTGAAAAAGACGAGTACACTCTGACATCGAATCAGAACGATCAGCTGGACG 796
DB      941 TGGGGCTGTGTCTGACTTCTACACACTCCTGTGGGATCCGCACTGTGGCTGTACCA 1000
QY      797 AGAAGAGGCTCTATCTGAACGGGAACTGTCTTTTGAAGGCTTTGAAAGCAGCAGG 856
DB      1001 AGAGCCAGTTCCTCATCAATGGGAAACCTTCTATTTCACGGGTGTCAACAAGCATGAG 1060
QY      857 AATTCCCGTTCTGGGACGGGACCTTTTATCCATTGATGATAAAAGCTTCAACCTTC 916
DB      1061 ATGCGGACATCCGAGGAGGAGGCTTCGACTGGCCGCTGCTGTGAAGACTTCAACCTGC 1120
QY      917 TGAAGTGATCAACGGAATTTCTTCAAGACCTCTCACTATCTTACAGTGAAGAGTGGC 976
```

```
DB      1121 TTGGCTGGCTTGGTGCCAACGCTTCCGTACCAAGCCACTACCCCTATGCAAGAGATGA 1180
QY      977 TGAATCTTCCGACAGACTCGAATCCTTGTGATAGACGAAGCCCGCAGCTTGGTATCA 1036
DB      1181 TGCAGATGTGTGACCGCTATGGATTTGTGTGTCATGATGAGTGTCCGGCGTGGTCTGG 1240
QY      1037 CAAGGTACCA---CTACATCCCGAGACTCAGAAGATAGCAGAAGACAACATAGAGAA 1093
DB      1241 CGTGCCGAGATTTCTTCAACAACGTTTCTGTGATCACCACATGAGGTGATGAGAGAA 1300
QY      1094 TGATCGACAGACAGAAAGAACCATCCAGTGTGATCATGTGAGTGTGCGGAACGAACAG 1153
DB      1301 TGGTGCGTAGGAGACAAGAACCAACCCCGGTGCTGTGATGTGTCTGTGGCCAACGAGCTG 1360
QY      1154 AGTCCAACCATCCAGACGCGGAGGTTTCTCAAGCCCTTATGAGACTGCCAATGAAA 1213
DB      1361 CGTCCACCTAGAACTGTGCTGGCTACTACTTGAAGATGTGATCGCTCACACCAATCCT 1420
QY      1214 TGATCGAACACGCCCCGTTGTGATGTGAGCATGATGAGCCGACGACGAGAGAACAA 1273
DB      1421 TGGACCCCTCCCGGCTGTGACCTTGTGAGCACTTAATACTATGAG----- 1468
QY      1274 GAGACGTGGCGCTGAAGTACTTTCGACATCGTCTGTGTAACAGGTACTACGGCTGTACA 1333
DB      1469 CAGACAAAGGGGCTCCGTATGTGATGTGATCTGTTGAACAGCTACTACTTGTGATC 1528
QY      1334 TCTATCAGGGAAGATAGAGAAGACTTCAAGCTCTGGAATAAGACATAGAGAGCTCT 1393
DB      1529 ACGACTACGGGACACTGAGTGTGATTCAGCTGCACTGGCCACCCAGTTTGAAGACGTGT 1588
QY      1394 ATGCAAGGCACAGAAAGCCCATCTTGTGCACAGAAATTCGGTCCGACCGGATAGCTGCA 1453
DB      1589 ATAGAAAGTATCAGAAAGCCCATTTTACAGAGCGAGTATGAGCAGAAACGATTGCGGGT 1648
QY      1454 TCCACTACGATCCACCTCAATGTTCTCCGAAGTACCAAGCAGAGCTCGTTGAAGA- 1512
DB      1649 TTCACCAAGATCCACCTCTGATGTTCACTGAAGAGTACCAAGAAAGTCTGCTAGACGAT 1708
QY      1513 ACGATCAGGCTCCTTTTGAAGAA----AGACTACATCATCGGAACACACGCTGTGGGCT 1567
DB      1709 ACCATCTGGGTCTGATCAAAAAACGAGAAATATGTGTTGAGAGCTCATTTGGAATT 1768
QY      1568 TTGCAGATTTTAAAGTCTCTCAGATGTGAGAAAGCCATTCTCAACCAAGGGTCTT 1627
DB      1769 TTGCCGATTTTCACTGACTGACAGATACCGACGAGAGTGTGGGGAATAAAAAAGGGATCT 1828
QY      1628 TCACAGAGACAGACAAACCCAACTGTTGCTCATGTACTGAGA 1671
DB      1829 TCACTCGCAGAGACAAACAAAAGTGCAAGCGTCTTTTGGCA 1872
XX      RESULT 6
XX      AAD06386
XX      ID AAD06386 standard; DNA; 1956 BP.
XX      AC AAD06386;
XX      DT 10-AUG-2001 (first entry)
XX      DE Human beta-glucuronidase (GUS) DNA.
XX      KW Human, adeno-associated viral expression vector; AAV; gene therapy;
XX      lysosomal storage disease; LSD; mucopolysaccharidoses VII; MPS VII;
XX      Sly syndrome; beta-glucuronidase; GUS; glycosaminoglycan; GAG; ds.
XX      OS Homo sapiens.
XX      FH Key Location/Qualifiers
XX      FT CDS 1..1956
XX      FT /tag= a
XX      FT /product= "Human beta-glucuronidase (GUS) "
```

PN WO200136603-A2.  
XX  
PD 25-MAY-2001.  
XX  
PF 17-NOV-2000; 2000WO-US031688.  
XX  
PR 17-NOV-1999; 99US-0166097P.  
PR 30-JUN-2000; 2000US-0215430P.  
XX  
PA (AVIG-) AVIGEN INC.  
PA (CHIL-) CHILDREN'S HOSPITAL MEDICAL CENT.  
PI Podsakoff G, Watson G, Couto LB, Yang B;  
XX  
XX WPI; 2001-343814/36.  
DR P-PSDB; AAE02443.  
XX  
PT Use of recombinant adeno-associated virus, comprising gene encoding a  
PT protein defective or missing in lysosomal storage disease, in the  
PT manufacture of a medicament for treating the lysosomal storage disease.  
XX  
PS Disclosure; Page 77-80; 97pp; English.  
XX  
CC The present invention relates to recombinant adeno-associated virus (AAV)  
CC expression vectors and virions, which include genes coding for enzymes  
CC defective or missing in lysosomal storage disease (LSD). AAV is useful in  
CC the manufacture of a medicament for treating lysosomal storage disease  
CC e.g., mucopolysaccharidoses VII (MPS VII). MPS VII (Sly Syndrome) is due  
CC to the deficiency in the lysosomal enzyme beta-glucuronidase (GUS) which  
CC aids in the breakdown of glycosaminoglycans (GAGs). AAV is used in gene  
CC therapy. The present sequence is human GUS DNA. This sequence is used in  
CC AAV constructs  
XX  
SQ Sequence 1956 BP; 437 A; 533 C; 563 G; 423 T; 0 U; 0 Other;

Query Match 10.2%; Score 172.4; DB 4; Length 1956;  
Best Local Similarity 52.1%; Pred. No. 7.5e-43;  
Matches 492; Conservative 0; Mismatches 431; Indels 21; Gaps 4;

QY 737 TGGAACTGAAAAAGACGAGTACTCTGGACATCGGAATCAGAACGATCAGCTGGGACG 796  
DB 938 TGGGGCCTGTGTCTGACTTCTACACACTCCCTGTGGGATCCGCACTGTGGCTGTCA 997  
QY 797 AGAAGAGGCTCTATCTGAACGGGAAACCTGTCTTTTGAAGGCTTTGAAAGCAGCAGG 856  
DB 998 AGAGCCAGTTCTTCATCAATGGGAAACCTTTCTATTTCACAGGTGTCAACAAGCATGAGG 1057  
QY 857 AATTCCCCGTTCTGGGGCAGGGCAGCTTTTATTCATGTATGATAAAAGACTTCAACCTTC 916  
DB 1058 ATGCGGACATCGAGGGAGGGCTTGACTGGCCGCTGCTGTGAAGAACTTCAACCTGC 1117  
QY 917 TGAAGTGGATCAACGCGAATTTTTCAGGACCTCTCACTATCTTACAGTGAAGTGGC 976  
DB 1118 TTGCGCTGGCTTGTGTGCCAACGCTTTCGTACCACTACCTCTATGACAGGAAGTGA 1177  
QY 977 TGGATCTTGCCGACAGACTCGGAATCCTTGTGATAGACGAAGCCCGCAGCTTGTATCA 1036  
DB 1178 TGCAGATGTGTACCGCTATGGATGTGTCAATGATGATGTGCCGCTGGCCCTGG 1237  
QY 1037 CAAGGTACCA---CTACAATCCGAGACTCAGAAGATGACAGAACAACATAAGAAGAA 1093  
DB 1238 CGCTGCCGACAGTTCTTCAACAACGTTTCTGTGCATCACCAATGACAGTGAAGAAG 1297  
QY 1094 TGATCGACAGACAGAAGAACCATCCAGTGTGATCATGTGAGTGTGGCGAAGAACAG 1153  
DB 1298 TGTGCGTAGAGGACAAGAACCAACCCCGGCTGTGTGTGTGTGGCCAAGACCTG 1357  
QY 1154 AGTCCAAACCATCCAGACGCGGAGGTTTCTTCAAGCCCTTTATGAGACTGCCAATGAAA 1213  
DB 1358 CGTCCACCTTAAGATCTGTGGCTACTACTTGAAGATGTGATCGCTCACACCAATCCT 1417  
QY 1214 TGGATCGAACACGCCCCGTTGTCAATGTGAGCATGTAGACGACACGACGAGAGAACA 1273

DB 1418 TGGACCCCTCCCGGCTGTGACTTTGTGAGCAACTTAATATGACG----- 1465  
QY 1274 GAGACGTGGCGCTGAAGTACTTGCACATCGTCTGTGTGAACAGTACTACGGCTGTACA 1333  
DB 1466 CAGACAAGGGGCTCCGTATGTGATGTGATGTGATCTGTTGAACAGTACTACTTGTGTATC 1525  
QY 1334 TCTATCAGGGAAGATTAAGAAGAGACTTCAAGCTCTGMAAAAGACATAGAGCTCT 1393  
DB 1526 ACGACTACGGGACCTGAGTTGATTCAGCTGCAGCTGGCCACCAGTTTGAACCTGCT 1585  
QY 1394 ATGCAAGCAGACAGAAAGCCCATCTTTGTACAGAAATTGGTGGGACGCGATAGCTGCA 1453  
DB 1586 ATAGAAGTATCAGAAAGCCCATTAATTCAGAGCGAGTATGAGCAGAAACGATTGCAGGT 1645  
QY 1454 TCCACTACGATCCACCTCAAAATGTTCTCCGAGAGTACCAGAGAGCTGTTGAAAAG- 1512  
DB 1646 TTCACACAGATCCACCTCTGATGTTCACTGAAGAGTACCAGAAAGTCTGTAGAGCAGT 1705  
QY 1513 ACGATCAGGCTCCTTTTGAAGAA-----AGACTACATCATCGGAACACAGTGGGCT 1567  
DB 1706 ACCATCTGGGTCTGATCAAAAACGAGAAATATGTGTGAGAGAGCTCATTTGGAATT 1765  
QY 1568 TTGCAGATTTTAAGACTCCTCAGATGTGAGAAAGACCATTCTCAACCAAGGCTGTTT 1627  
DB 1766 TTGCCGATTTGATGACTGAACAGTCAACGACAGAGAGTGTGGGAATAAAAGGGATCT 1825  
QY 1628 TCACAAGAGACAGACCAACCAAACTGTTGCTCATGTACTAGAA 1671  
DB 1826 TCACCTCGCAGAGACACCAAAAAGTGACGCTTCTTTTGCA 1869

RESULT 7  
ABK83932  
ID ABK83932 standard; cDNA; 2191 BP.  
XX  
AC ABK83932;  
XX  
DT 14-AUG-2002 (first entry)  
XX  
DE Human cDNA differentially expressed in granulocytic cells #503:  
XX  
KW Human; ss; granulocytic cell; DNA chip; bacterial infection;  
KW viral infection; parasitic infection; protozoal infection;  
KW fungal infection; sterile inflammatory disease; psoriasis;  
KW rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;  
KW cardiac reperfusion injury; renal reperfusion injury; ARDS;  
KW adult respiratory distress syndrome; inflammatory bowel disease;  
KW Crohn's disease; ulcerative colitis; periodontal disease;  
KW granulocyte activation; chronic inflammation; allergy.  
OS Homo sapiens.  
XX  
PN WO200228999-A2.  
XX  
PD 11-APR-2002.  
XX  
PF 03-OCT-2001; 2001WO-US030821.  
XX  
PR 03-OCT-2000; 2000US-0237189P.  
XX  
PA (GENE-) GENE LOGIC INC.  
XX  
PI Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;  
XX  
DR WPI; 2002-435328/46.  
XX  
PT Detecting granulocyte activation by detecting differential expression of  
PT genes associated with granulocyte activation, which serves as diagnostic  
PT markers that is useful for monitoring disease states and drug toxicity.  
XX  
PS Claim 1; SEQ ID NO 503; 114pp; English.  
XX  
CC The invention relates to detecting (M1) granulocyte (GC) activation

(GCA), by detecting the level of expression of gene(s) (Gs) identified by DNA chip analysis as given in the specification, and comparing the expression level to an expression level in an unactivated GC, where differential expression of Gs is indicative of GCA. Also included are modulating (M2) GA by contacting GC with an agent that alters the expression of at least one gene in Gs; (2) screening (M3) for an agent capable of modulating GCA or an inflammation (especially chronic) in a tissue, an allergic response in a subject, exposure of a subject to a pathogen or sterile inflammatory disease using the gene expression profile; (3) detecting (M4) an inflammation (especially chronic) in a tissue, an allergic response in a subject, exposure of a subject to a pathogen or sterile inflammatory disease, by detecting the level of expression in a sample of the tissue of gene(s) from Gs, where the level of expression of the gene is indicative of inflammation; (4) treating (M5) an inflammation (especially chronic) or in a tissue, an allergic response in a subject, exposure of a subject to a pathogen or sterile inflammatory disease, by contacting a tissue having inflammation with an agent that modulates the expression of gene(s) from Gs in the tissue. M1 is useful for detecting GCA; M2 is useful for modulating GA; M3 is useful for screening an agent capable of modulating GCA preferably in an inflammation in a tissue; M4 is useful for detecting an inflammation (especially chronic) in a tissue, an allergic response in a subject, exposure of a subject to a pathogen or sterile inflammatory disease, by contacting a tissue having inflammation with an agent that modulates the expression of gene(s) from Gs in the tissue. M5 is useful for detecting GCA; M2 is useful for modulating GA; M3 is useful for screening an agent capable of modulating GCA preferably in an inflammation in a tissue; M4 is useful for detecting an inflammation (especially chronic) in a tissue, an allergic response in a subject, exposure of a subject to a pathogen or sterile inflammatory disease (e.g. psoriasis, rheumatoid arthritis, glomerulonephritis, asthma, thrombosis, cardiac reperfusion injury, renal reperfusion injury, ARDS, adult respiratory distress syndrome, inflammatory bowel disease, Crohn's disease, ulcerative colitis, periodontal disease; also bacterial infection, viral infection, parasitic infection, protozoal infection, fungal infection and M5 is useful for treating one of the above conditions. The present sequence represents a gene differentially expressed in granulocytes. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

Sequence 2191 BP, 497 A; 589 C; 630 G; 475 T; 0 U; 0 Other;

Query Match 10.2%; Score 172.4; DB 6; Length 2191;  
Best Local Similarity 52.1%; Pred. No. 8e-43;  
Matches 492; Conservative 0; Mismatches 431; Indels 21; Gaps 4;

OY 737 TGAAGTGAAGAAAGACGAGTACCTTGACATCGGAATCGAAGCATGAGCTGGAGC 796  
DB 964 TGGGGCTGTGTCTGACTTCTACACACTCCCTGTGGGATCCGCACTGTGGCTGTACCA 1023  
OY 797 AGAAGAGGCTCTATCTGAACGGGAAACCTGCTTTTGAAGGGCTTTGAAAGCAGCAGG 856  
DB 1024 AGAGCCAGTTCCTCATCAATGGGAAACCTTCTATTCCACGGTGTCAACAAGCATGAGG 1083  
OY 857 AATTCCCGCTTCTGGGACGAGGACCTTTTATCCATTGATGATTAAGACTTCAACCTTC 916  
DB 1084 ATGCGGACATCCGAGGGAAGGGCTTGAAGTGGCCGCTGTGTGAAGGACTTCAACCTGC 1143  
OY 917 TGAAGTGAATCAACGCGAATCTTTCAGAGCCTCTCACTATCTTACAGTGAAGAGTGGC 976  
DB 1144 TTGCTGCTGTGTGTCACACGCTTTCGTAACGACCACTACCCCTATGCAAGAGAGTGA 1203  
OY 977 TGGATCTTGGCCGACAGACTCGGAATCTTGTGATAGACGAGCCCGCAGCTGTGTATCA 1036  
DB 1204 TGCAGATGTGACCGCTATGGGATTTGTGTCATGATGATGTCCCGCGTGGCCCTGG 1263  
OY 1037 CAAGGTACCA--CTACAATCCGAGACTCAGAAAGATAGAGAAACAATGAAGAA 1093  
DB 1264 CGCTGCCGAGTTCTTCAACAAGTTTCTTGATCACCACATGAGGTGATGGAAGAG 1323  
OY 1094 TGATGACAGACACAAGAACCATCCAGTGTGATCATGTGAGTGTGGCGAAGAACAG 1153  
DB 1324 TGGTGCAGTGGGACAAAGAACCAACCCCGCGTGTGATGTGTGTGCGCAACGAGCCTG 1383  
OY 1154 AGTCAACCATTCAGACGCGGAGGTTTCTTCAAGCCCTTATGAGACTGCCAATGAAA 1213  
DB 1384 CGTCCACCTAGAAATCTGTGGCTACTACTGAAGATGTGATCGCTACACCAATCTCT 1443

OY 1214 TGGATCGAACACGCCCCGTTGTCTATGTGTAGCATGATGACCGACCAAGCAGAGAACAA 1273  
DB 1444 TGAACCCCTCCGGGCTGTGACTTGTGTAGCACTTAATCTATGCAAG----- 1491  
OY 1274 GAGACGTGGCGCTGAAGTACTTTCGACATCGTCTGTGTGAACAGGTAAGCTGCTGTA 1333  
DB 1492 CAGACAAGGGGGCTCCGTATGTGATGTGATGTGATCTTTGAACAGCTACTCTTGTGATC 1551  
OY 1334 TCTATCAGGGAAGATAGAAAGGACTTCAAGCTCTGGAAGAAACATAGAGAGCTCT 1393  
DB 1552 ACGACTACGGGCACTGTGAGTTGATTCAGCTGAGCTGGCCACCAAGTTGAGAACTGGT 1611  
OY 1394 ATGCAAGGCACAGAAAGCCCATCTTTGTCAACAGATTCGGTGGGAGCGGATAGCTGCA 1453  
DB 1612 ATAGAAATATCAGAAAGCCCATTTATTCAGAGCGAGTATGAGCAGAAACGATTGCAAGGT 1671  
OY 1454 TCCACTACGATCCACTCAAAATGTTCTCCGAAAGATACCAAGCAGAGCTCGTTGAAAAG- 1512  
DB 1672 TTGACACAGATCCACTGTGATGTTCAGTGAAGAGTACCAAGAAAGTCTGCTAGACAGT 1731  
OY 1513 ACGATCAGGCTCTTTGAAAAA-----AGACTACATCATCTCGGAACACAGCTGTGGGCT 1567  
DB 1732 ACCATCTGGGTCTGATCAGTCAAAACGAGAAATATGTGTGGAGAGCTCATTTGGAATT 1791  
OY 1568 TTGCAAGATTTTAAGACTCTCTCAGATGTGAGAAAGCCCATTTCAACCAAGGCTGTT 1627  
DB 1792 TTGCGATTTCAATGATCAAGACAGTCAACGAGAGAGTCTGGGAAATAAAGGGATCT 1851  
OY 1628 TCACAAGAGACAGCAACCCAACTCGTTGCTCATGTACTGAGA 1671  
DB 1852 TCACTCGGACAGACAAACCAAAAGTGACGCTTCTTTGCCA 1895

RESULT 8

ADD29602  
ID ADD29602 standard; mRNA; 2191 BP.

AC ADD29602;

DT 15-JAN-2004 (first entry)

DE Human tumour suppressor mRNA SEQ ID NO:55.

OS ss; human; tumour suppressor; cancer; cancer; cytosolic; gene therapy.

XX Homo sapiens.

PN WO2003058201-A2.

PD 17-JUL-2003.

PF 31-DEC-2002; 2002WO-US041825.

PR 31-DEC-2001; 2001US-0345317P.

PA (QUAR-) QUARK BIOTECH INC.  
(CLEV-) CLEVELAND CLINIC FOUND.

PI Feinstein E, Gudkov AV;

DR WPI; 2003-598393/56.

Diagnosing cancer comprises determining the polypeptide or polynucleotide levels e.g., hepatic lipase, in a sample from a subject, where a higher level compared to that in a subject free of cancer is indicative of cancer.

Disclosure; SEQ ID NO 55; 272bp; English.

The invention relates to a novel method for diagnosing a cancer in a subject. the method comprises determining, in a sample from the subject, the level of at least one polypeptide, where a higher level of the polypeptide compared to the level of the polypeptide in a subject free of





Db 964 TGGGGCCTGTCTGACTTCTACACACTCCCTGTGGGATCCGCACTGTGCTGTACCA 1023  
QY 797 AGAAGAGCTCTATCTGAAGGGAACCTGTCTTTTGAAGGCTTTGAAGCAGCAGG 856  
Db 1024 AGAGCCAGTTCCTCATCAATGGGAAACCTTCTATTCCAGGCTGTCAACAAGCATGAGG 1083  
QY 857 AATTCCCGCTTCTGGGCGAGGGCACCTTTTATCCATTGATGATAAAGACTTCAACCTTC 916  
Db 1084 ATGCGGACATCCGAGGGAAGGGCTTCGACTGGCCGCTGCTGTGAAGACTTCAACCTGC 1143  
QY 917 TGAAGTGATCAACGCGAATCTTTCAGGACTCTCACTATCCTTACAGTGAAGAGTGGC 976  
Db 1144 TTCGCTGGCTTGGTGGCAACGCTTCCGTACAGCCACTACCCCTATGAGAGAGTGA 1203  
QY 977 TGGATCTTGGCGACAGACTCGGAATCCTTGTGATAGACGAAGCCCGCAGCTTGTATCA 1036  
Db 1204 TGCAGATGTGTGACCGCTATGGGATGTGTGCATCGATGATGTCCCGGCTGGCCTGG 1263  
QY 1037 CAAGGTACCA---CTACAATCCCGAGACTCAGAGATAGCAGAAGACAACATAAGAGAA 1093  
Db 1264 CGCTGCCGCAGTCTTCAACAACGTTCTCTGCATCACCACATGAGGTGATGGAAGAG 1323  
QY 1094 TGATCGACAGACAGAAGAACATCCAGTGTGATCATGTGAGTGTGGCGAAGCAACAG 1153  
Db 1324 TGGTGCCTAGGAGACAAGAACACCCCGCGGTGTGATGTGTGTGGCCACAGACCTG 1383  
QY 1154 AGTCCCAACCATCCAGACCGGAGGGTCTTCAAGCCCTTATGAGACTGCCAATGAAA 1213  
Db 1384 CGTCCCAACCATAGATCTGTGGCTACTACTTGAAGATGTGATCGCTCACACCAATCCT 1443  
QY 1214 TGGATCGAAGACGCCCCGCTGTCTGATGTGAGCATGATGAGCAGCAGCAGAGAGAA 1273  
Db 1444 TGGACCCCTCCCGGCTGTGACCTTGTGAGCAACTTAACTATGACAG----- 1491  
QY 1274 GAGACGTGGCGCTGAAGTACTTGCACATCGTCTGTGTAACAGGTACTACGGCTGTACA 1333  
Db 1492 CAGACAGGGGGCTCCGTATGTGATGTGATCTGTGTAACAGTACTACTTGTGTATC 1551  
QY 1334 TCTATCAGGGAAGATAGAGAAGACTTCAAGCTCTGGAAGAAAGACATGAGAAGCTCT 1393  
Db 1552 ACGACTACGGGACCTGAGTGTGATTCAGTGCAGCTGCGCACCCAGTTGAGAAGTGT 1611  
QY 1394 ATGCAAGGCAAGAAAGCCACTTGTGTCAAGAAATTCGGTGGCGGAGCGCATAGCTGCA 1453  
Db 1612 ATAAAGATATCAGAAAGCCACTTATTCAGAGCGAGTATGAGACAGAAAGATTCAGGGT 1671  
QY 1454 TCCACTACGATCCACTCAATGTCTCCGAAGAGTACCAAGCAGAGCTCGTTGAAAAG- 1512  
Db 1672 TTCAACGAGATCCACTGTGATGTTCAGTAAAGATACCAAGAAAGTCTGCTAGAGCAGT 1731  
QY 1513 ACGATCAGGCTCTTTGAAAAA-----AGACTACATCATCGAACAACAGTGTGGGCT 1567  
Db 1732 ACCATCTGGGTCTGATCAAAAACGAGAAATATGTGTTGAGAGCTCATTTGGAATT 1791  
QY 1568 TTGAGATTTTAAGACTCCTCAGATGTGAGAAGACCAATTCTCAACCAAGAGGTGTTT 1627  
Db 1792 TTGCCGATTCATGACTGAACAGTCAACGAGAGAGTGTGGGGAATAAAAAGGGGATCT 1851  
QY 1628 TCACAAGAGACAGAACCAAACTCGTGTCTCATGTACTGAGA 1671  
Db 1852 TCACCTCGGAGAGACAAACCAAAAGTGCAGCGTTCCTTTGCGA 1895

RESULT 10  
ACF57506  
ID ACF57506 standard; DNA; 2191 BP.

AC ACF57506;  
XX  
DT 22-APR-2004 (first entry)  
XX  
DE Human beta-glucuronidase DNA SEQ ID NO: 7.

XX  
KW Protein targeting; angiogenesis inhibitor; fusion gene; cancer; gene; ds;  
KW antiangiogenic; cyostatic; ophthalmological; antiarthritic;  
KW antineumatic; antiinflammatory; dermatological; immunosuppressive;  
KW antipsoriatic; antidiabetic; neuroprotective; vasotropic; anorectic.  
OS Homo sapiens.  
XX  
PN WO2003093303-A1.  
XX  
PD 13-NOV-2003.  
XX  
PF 06-MAY-2003; 2003WO-US014243.  
XX  
PR 06-MAY-2002; 2002US-0380063P.  
XX  
PA (TEXA ) UNIV TEXAS SYSTEM.  
XX  
PI Hung M, Ian K, Ou-Yang F, Liu J, Ian K;  
XX WPI; 2004-053038/05.  
XX  
PT New polypeptide comprising an antiangiogenesis polypeptide region linked  
PT to a therapeutic protein or a diagnostic protein, useful for diagnosing  
PT or treating angiogenesis-dependent diseases e.g. cancer, lupus or  
PT restenosis.  
XX  
PS Disclosure; Page 86; Opp; English.  
XX  
CC The present invention relates to a polypeptide comprising an  
CC antiangiogenesis polypeptide region linked to a therapeutic protein or  
CC polypeptide region, or a diagnostic protein or polypeptide region. The  
CC polypeptide comprising an antiangiogenesis polypeptide region or the  
CC nucleic acid encoding the polypeptide comprising the antiangiogenesis  
CC polypeptide is useful for diagnosing or treating angiogenesis-dependent  
CC diseases or cancer. Angiogenesis-dependent diseases include age-related  
CC macular degeneration, atherosclerosis, angiodioma, neovascular  
CC glaucoma, arteriovenous malformations, nonunion fractures, arthritis,  
CC rheumatoid arthritis, lupus, connective tissue disorders, Osler-Weber  
CC syndrome, psoriasis, corneal graft neovascularisation, pyogenic  
CC granuloma, delayed wound healing, retrolental fibroplasia, diabetic  
CC retinopathy, scleroderma, granulations, haemangioma, trachoma,  
CC haemophilic joints, vascular adhesions, hypertrophic scars, multiple  
CC sclerosis, restenosis or obesity. The present sequence is a coding  
CC sequence shown in the exemplification of the invention  
XX  
SQ Sequence 2191 BP; 497 A; 589 C; 630 G; 475 T; 0 U; 0 Other;  
Query Match 10.2%; Score 172.4; DB 12; Length 2191;  
Best Local Similarity 52.1%; Pred. No. 8e-43;  
Matches 492; Conservative 0; Mismatches 431; Indels 21; Gaps 4;  
QY 737 TGAAGCTGAAAAAGACGAGTACCTGTGACATCGGAATCAGAACGATCAGCTGGACG 796  
Db 964 TGGGCTGTGTCTGACTTCTACACACTCCCTGTGGGATCCGCACTGTGCTGTACCA 1023  
QY 797 AGAAGAGCTCTATCTGAACGGAACCTGTCTTTTGAAGGCTTTGAAAGCAGAGG 856  
Db 1024 AGAGCCAGTTCCTCATCAATGGGAAACCTTCTATTCCACGCTGTCAACAAGCATGAGG 1083  
QY 857 AATTCCCGCTTCTGGGCGAGGGCACCTTTTATCCATTGATGATAAAGACTTCAACCTTC 916  
Db 1084 ATGCGGACATCCGAGGGAAGGGCTTCGACTGGCCGCTGCTGTGAAGAGACTTCAACCTGC 1143  
QY 917 TGAAGTGATCAACGCGAATCTTTCAGGACTCTCACTATCCTTACAGTGAAGAGTGGC 976  
Db 1144 TTCGCTGGCTTGGTGGCAACGCTTCCGTACAGCCACTACCCCTATGAGAGAGTGA 1203  
QY 977 TGGATCTTGGCGACAGACTCGGAATCCTTGTGATAGACGAAGCCCGCAGCTTGTATCA 1036  
Db 1204 TGCAGATGTGTGACCGCTATGGGATGTGTGCATCGATGATGTCCCGGCTGGCCTGG 1263  
QY 1037 CAAGGTACCA---CTACAATCCCGAGACTCAGAAAGATAGCAGAAGACAACATAAGAGAA 1093

Db 1264 CGCTGCCGAGTTCTTCAACAACGTTTCTCTGCATCACACATGACGGTGTAGGAAGAAG 1323  
QY 1094 TGATCGACAGACACAGAAGCAATCCAGTGTGATCATGTGAGTGTGGCAAGCAACGAG 1153  
Db 1324 TGTGTCGTAGGACAGAAACCAACCCGCGTGTGATGTGTCTGTGGCCAAACGAGCCTG 1383  
QY 1154 AGTCCAAACCATCCAGACGCGGAGGGTTCTTCAAAAGCCTTTATGAGACTGCCAATGAAA 1213  
Db 1384 CGTCCCACTAGAAATCTGCTGCTACTACTTGAAGATGGTGTATCGCTCACACCAATCCT 1443  
QY 1214 TGGATCGAACACGCCCCGTTGTGATGTGAGCATGATGAGACGACACGAGAGAACAA 1273  
Db 1444 TGGACCCCTCCCGGCTGTGACCTTGTGAGCAACTTAATACTATGAG----- 1491  
QY 1274 GAGACGTGGCGCTGAGTACTTGCATCGTCTGTGGAACAGGTACTACGGCTGGTACA 1333  
Db 1492 CAGACAAAGGGGCTCCGTATGTGATGTGATCTGTTGAACAGCTACTACTCTTGGTATC 1551  
QY 1334 TCTATCAGGGAAGATAGAGAAGACTTCAAGCTCTGGAAGAAAGACATAGAAGCTCT 1393  
Db 1552 ACGACTACGGGACCTGAGTGTGATTCAGCTGCAGCTGGCCACCCAGTTTGAGAACTGGT 1611  
QY 1394 ATGCAAGGCACAGAAAGCCCATCTTGTGCACAGAAATTCGGTGGCGACGGCATAGCTGCA 1453  
Db 1612 ATAGAAGATATCAGAAAGCCCATTTATCAGAGCGAGTATGAGACAGAAAGCATTTGCAAGGT 1671  
QY 1454 TCCACTACGATCCACTCAATGTCTCCGAAGATACCAAGACAGCTCGTTGAAAAG- 1512  
Db 1672 TTCACCAAGATCCACTCTGATGTTCATGAAAGATACCAAGAAAGTCTGCTAGACAGT 1731  
QY 1513 ACGATCAGGCTCCTTTTGAAAAA-----AGACTACATCATCGGAACACACGTTGGGCT 1567  
Db 1732 ACCATCTGGGTCTGATCAAAAAACGCAAGAAATATGTGTTGAGAGCTCATTTGGAATT 1791  
QY 1568 TTGCAGATTTTAAGACTCCTCAGAAATGTGAGAAAGCCATTCTCAACCAAGGGTGT 1627  
Db 1792 TTGCCGATTTTCATGACTGAAACAGTACCGACGAGAGTCTGGGGAATAAAAAGGGGATCT 1851  
QY 1628 TCACAAGAGACAGCAACCCAACTGTTGCTCATGTACTGAGA 1671  
Db 1852 TCACTCGGACAGACCAACCAAAAGTGCAAGCGTCTTTTGGCA 1895

RESULT 11  
ADJ62864  
ID ADJ62864 standard; DNA; 2191 BP.

AC ADJ62864;  
DT 06-MAY-2004 (first entry)  
XX Human beta-glucuronidase gene related to cancer treatment.  
DE  
XX beta-catenin; bipartite T-cell factor; Tcf; promoter construct; LEF-1;  
KW cyostatic; beta-catenin activity inhibitor; gene therapy; colon cancer;  
KW metastasis; liver; thymidine kinase; prodru; chemotherapy;  
KW radiation therapy; surgery; human; beta-glucuronidase; gene; ds.  
XX  
OS Homo sapiens.  
XX  
PN US2003228285-A1.  
XX  
PD 11-DEC-2003.  
XX  
PF 05-MAY-2003; 2003US-00429802.  
XX  
PR 03-MAY-2002; 2002US-0377672P.  
XX  
PA (HUNG/) HUNG M.  
PA (KMON/) KWONG K Y.  
PA (ZOU/) ZOU Y.  
XX

PI Hung M, Kwong KY, Zou Y;  
XX  
DR WPI; 2004-042209/04.  
XX  
PT Novel viral vector comprising beta-catenin/bipartite T-cell factor-  
PT responsive promoter having first and second promoter region linked to  
PT target nucleic acid sequence, useful for treat- ing colon cancer.  
XX  
PS Disclosure; SEQ ID NO 16; 114pp; English.  
XX  
CC This invention relates to a novel viral vector comprising a beta-  
CC catenin/bipartite T-cell factor (Tcf)-responsive promoter construct which  
CC contains a first promoter region having a copy of Tcf/LEF-1 binding site  
CC operatively linked to a second promoter region, and a nucleic acid  
CC sequence, where the first and second promoter regions are operatively  
CC linked to target nucleic acid sequence. The invention may be useful for  
CC the development of compounds with a cyostatic activity, through action  
CC as beta-catenin activity inhibitors, or for gene therapy. The invention  
CC may be useful for treating an individual with colon cancer which is  
CC metastasised to the liver. The treatment involves administering the  
CC vector of the invention where the nucleic acid sequence encodes a  
CC therapeutic polypeptide or thymidine kinase, a prodru and chemotherapy,  
CC radiation, surgery or gene therapy to the individual. The present  
CC sequence is that of a therapeutic gene which may be used in the vector of  
CC the invention for the treatment of an individual with cancer.  
XX  
SQ Sequence 2191 BP; 497 A; 589 C; 630 G; 475 T; 0 U; 0 Other;

Query Match 10.2%; Score 172.4; DB 12; Length 2191;  
Best Local Similarity 52.1%; Pred. No. 8e-43;  
Matches 492; Conservative 0; Mismatches 431; Indels 21; Gaps 4;

QY 737 TGAAGCTTGAAGAAAGACGATACCTCTGACATCGGAATCAGAACGATCAGCTGGAGC 796  
Db 964 TGGGGCTGTGTCTGACTTCTACACACTCCCTGTGGGATCCGCACTGTGCTGTACCA 1023  
QY 797 AGAAGAGCTTATCTGAACGGGAACCTGTCTTTTGAAGGCTTTGAAGACGAGG 856  
Db 1024 AGAGCCAGTTCCTCATCAATGGGAAACCTTTCTAATTTCCACGGTGTCAACAGCATGAGG 1083  
QY 857 AATTCCCCCTTCTGGGGCAGGGACCTTTATCCATTGATGATATAAAGACTTCAACCTTC 916  
Db 1084 ATCCGACATCCGAGGGAAGGGCTTCGACTGGCGCTGTGGAAGACTTCAACCTGC 1143  
QY 917 TGAAGTGATCAACGGGAATTTCTCAGGACCTCTCATCTATCTTACAGTGAAGAGTGGC 976  
Db 1144 TTGGCTGGCTGTGTCCAACGCTTTCCGTACCAGCCACTATCCCTATGACAGGAAGTGA 1203  
QY 977 TGGATCTTGGCAGACAGACTCGGAATCCTTGTGATAGACGAAGCCCCGACGTTGGTATCA 1036  
Db 1204 TGCAGATGTGACCGCTATATGGGATGTGTGATCATGATGATGTCCCGGCGTGGCCTGG 1263  
QY 1037 CAAGGTACCA---CTACAATCCCGAGACTCAGAAAGATAGCAGAAAGCAACATTAAGAA 1093  
Db 1264 CGCTGCCGAGTTCTTCAACAACGTTTCTGTGATCACCAATGCAAGGTAGGAAGAAAG 1323  
QY 1094 TGATCGACAGACAGAAAGCAATCCAGTGTGATCATGTGAGTGTGGCAACGAACGAG 1153  
Db 1324 TGTGTCGTAGGACAGAACCAACCCGCGGTGTGATGTGTGTGGCCAAAGACCTG 1383  
QY 1154 AGTCCAAACCATCCAGACGCGGAGGGTTCTTCAAAAGCCCTTTATGAGACTGCCAATGAAA 1213  
Db 1384 GTCCCACTAGAAATCTGTGCTACTACTTGAAGATGTGTGCTCACACCAATCCT 1443  
QY 1214 TGGATCGAACACGCCCCGTTGTGATGTGAGCATGATGAGCGACACGAGAGAAACAA 1273  
Db 1444 TGGACCCCTCCCGGCTGTGACCTTGTGAGCAACTTAATACTATGAG----- 1491  
QY 1274 GAGACGTGGCGCTGAGTACTTGCATCGTCTGTGTAACAGGTACTACGGCTGTACA 1333  
Db 1492 CAGACAAAGGGGCTCCGTATGTGATGTGATCTGTTGAACAGCTACTACTTGTGATC 1551  
QY 1334 TCTATCAGGGAAGATAGAGAAGGACTTCAAGCTCTGAAAAAAGACATAGAAGAGCTCT 1393

```
DB 1552 ACGACTACGGGACCTGAGTGTGATTCAGCTGCAGCTGGCCACCCAGTTTGAGAACTGGT 1611
QY 1394 ATGCAAGGCACAGAAAGCCCATCTTTGTTCACAGATTGCGTGGGACCGGATAGCTGGCA 1453
DB 1612 ATAGAAATATCAGAAAGCCCATTTATTCAGAGCGAGTATGAGCAGAAACGATTGCGAGGT 1671
QY 1454 TCCACTACGATCCACCTCAAAATGTTCTCCGAAAGATACCAAGCAGAGCTCGTTGAAAG- 1512
DB 1672 TTCACGAGGATCCACCTCTGATGTTCACTGAAGATACCAAGAAAGTCTGCTAGACAGT 1731
QY 1513 ACGATCAGGCTCCTTTGAAAAA-----AGACTACATCATCGGAACACAGTGTGGGCT 1567
DB 1732 ACCATCTGGGTCTGATCAAAAAACGCAAAATATGTGTTGAGAGCTCATTGGAAAT 1791
QY 1568 TTGCAGATTTTAAGACTCTCAGAAATGTGAGAGACCCATTCTCAACCAACAGGGTGT 1627
DB 1792 TTGCCGATTTTCATGACTGAACAGTCAACCGACGAGAGTGTGGGAAATAAAAAGGGGATCT 1851
QY 1628 TCACAAAGAGACAGACAAACCAAACTCGTTGCTCATGTAAGT 1671
DB 1852 TCACCTGGCAGAGACAAACCAAAAGTGACGCTTCTTTGGCA 1895
```

## RESULT 12

ADP10326  
ID ADP10326 standard; DNA; 2191 BP.

AC ADP10326;

DT 12-AUG-2004 (first entry)

DE Reference mRNA sequences for marker probe #3.

KM transplant rejection; immune system; rheumatoid arthritis; lupus;  
KW inflammatory bowel disease; multiple sclerosis; HIV; AIDS; ss.

OS Homo sapiens.

PN W02004042346-A2.

PD 21-MAY-2004.

PF 24-APR-2003; 2003WC-US012946.

PR 24-APR-2002; 2002US-00131831.

PR 20-DEC-2002; 2002US-00325899.

PA (EXPR-) EXPRESSION DIAGNOSTICS INC.

PI Wohlgenuth J, Fry K, Woodward R, Ly N, Prentice J, Morris M;  
PI Rosenberg S;

DR WPI; 2004-400724/37.

PT Diagnosing or monitoring transplant rejection, e.g. heart, kidney, liver,  
PT pancreas, pancreatic islet, lung, bone marrow or stem cell transplant  
PT rejection, in an individual, comprises detecting the expression level of  
PT the genes.

PS Claim 80; SEQ ID NO 335; 1762pp; English.

XX The present invention relates to diagnosing or monitoring transplant  
CC rejection, e.g. cardiac or kidney transplant rejection, in an individual  
CC comprises detecting the expression level of one or more genes. The  
CC methods, system and kits are useful in diagnosing or monitoring  
CC transplant rejection, e.g. heart, kidney, liver, pancreas, pancreatic  
CC islet, lung, bone marrow or stem cell transplant rejection,  
CC xenotransplant rejection or mechanical organ replacement rejection, in an  
CC individual. The method is also useful in assessing the immune status of  
CC an individual. The method is also useful in diagnosing and monitoring  
CC diseases that involve the immune system, e.g. rheumatoid arthritis,  
CC lupus, inflammatory bowel diseases, multiple sclerosis, HIV/AIDS or

CC viral, bacterial or fungal infection. The present sequence represents a  
CC mRNA reference sequence for a 50 mer oligonucleotide marker for diagnosis  
CC and monitoring of allograft rejection and other disorders.

XX Sequence 2191 BP, 497 A; 589 C; 630 G; 475 T; 0 U; 0 Other;

Query Match 10.2%; Score 172.4; DB 12; Length 2191;

Best Local Similarity 52.1%; Pred. No. 8e-43;  
Matches 492; Conservative 0; Mismatches 431; Indels 21; Gaps 4;

```
QY 737 TGAACCTTGAAGAAAGAGAGTACACTCTGACATCGGAATCAGAACGATCAGCTGGAGC 796
DB 964 TGGGGCTGTGTCTGACTTCTAACAACACTCCCTGTGGGATCCGACTGTGCTCACCA 1023
QY 797 AGAAGAGGCTCTATCTGAACGGGAAACCTGTCTTTTGAAGGCTTTGAAAGCAGAGG 856
DB 1024 AGAGCCAGTTCTCATCAATGGGAAACCTTTCTATTTCACGGTGTCAACAACATGAGG 1083
QY 857 AATTCCCGCTTCTGGGGCAGGGCACCCTTTTATCCATTGATGATTAAGACTTCAACCTTC 916
DB 1084 ATGGGACATCCGAGGAGGAGGGCTTCGACTGGCCGCTGTGTGAAGACTTCAACCTGC 1143
QY 917 TGAAGTGATCAACCGGAATCTTTCAGGACCTCTCACTATCTTACAGTGAAGAGTGGC 976
DB 1144 TTGCTGTGCTTGTGTGCCAACGCTTTCCGTACCAAGCACTACCCCTATGACAGAAAGTGA 1203
QY 977 TGGATCTTGGCCACAGACTCGGAATCCTTGTGATAGACGAAGCCCGCAGCTGTATCA 1036
DB 1204 TGCAGATGTGTGACCGCTATGGAGTTGTGTGATGATGATGATGATGATGATGATGATG 1263
QY 1037 CAAGGTACCA---CTCAATCCCGAGAGCTCAGAAAGATAGCAGAAAGACACATAGAAGA 1093
DB 1264 CGCTGCCGAGTTCTTCAACAACGTTTCTCTGATCACCACATGACGTGATGAAGAAG 1323
QY 1094 TGATCGACAGACAGCAAGAACCATCCAGTGTGATCATGTGAGTGTGGGAGCAACCAAG 1153
DB 1324 TGTGCGTAGAGGACAGAAACCAACCCCGGCTGTGATGTGTGTGTGTGTGTGTGTGTGTGT 1383
QY 1154 AGTCCAAACCATCCAGACGGGAGGTTTCTTCAAGCCCTTTATGAGACTGCCAATGAAA 1213
DB 1384 CGTCCCACTAGAAATCTGTGCTGCTACTACTTGAAGATGTGATGCTCACACCAATCTCT 1443
QY 1214 TGGATCGAACACGCCCCGTTGTCTATGTGTGAGCATGTGAGCCGACCAAGAGAAACA 1273
DB 1444 TGAACCTCTCCCGGCTGTGACCTTTGTGAGCACTCTAATCTATGACG----- 1491
QY 1274 GAGACGTGGCGCTGAAGTACTTGCACATCGTGTGTGTAACAGGTAATAAGGCTGTACA 1333
DB 1492 CAGACAAAGGGGCTCCGTATGTGATGTGATCTGTTTGAACAGCTACTACTCTGTGATC 1551
QY 1334 TCTATCAGGGGAAGATAGAAAGGACTTCAAGCTCTGAAAAAGACATAGAAAGCTCT 1393
DB 1552 ACGACTACGGGACACCTGAGTGAATTCAGCTGCAGCTGGCCACCAAGTTTGAAGACTGT 1611
QY 1394 ATGCAAGGCACAGAAAGCCCATCTTTGTACAGAAATTCGGTGGGACCGGATAGCTGGCA 1453
DB 1612 ATAGAAATATCAGAAAGCCCATTTATTCAGAGCGAGTATGAGCAGAAACGATTGCGAGGT 1671
QY 1454 TCCACTACGATCCACCTCAAAATGTTCTCCGAAAGATACCAAGCAGAGCTCGTTGAAAG- 1512
DB 1672 TTCACGAGGATCCACCTCTGATGTTCACTGAAGATACCAAGAAAGTCTGCTAGACAGT 1731
QY 1513 ACGATCAGGCTCCTTTGAAAAA-----AGACTACATCATCGGAACACAGTGTGGGCT 1567
DB 1732 ACCATCTGGGTCTGATCAAAAAACGCAAAATATGTGTTGAGAGCTCATTGGAAAT 1791
QY 1568 TTGCAGATTTTAAGACTCTCAGAAATGTGAGAGACCCATTCTCAACCAACAGGGTGT 1627
DB 1792 TTGCCGATTTTCATGACTGAACAGTCAACCGACGAGAGTGTGGGAAATAAAAAGGGGATCT 1851
QY 1628 TCACAAAGAGACAGACAAACCAAACTCGTTGCTCATGTAAGT 1671
DB 1852 TCACCTGGCAGAGACAAACCAAAAGTGACGCTTCTTTGGCA 1895
```





PR 17-SEP-1999; 99US-00399079.  
XX (GENZ ) GENZYME TRANSGENICS CORP.  
XX Pollock D, Meade HM, Bosslet K;  
XX WPI; 2001-257871/26.  
DR P-PSDB; AAB62267, AAB62268, AAB62269, AAB62270, AAB62271, AAB62272.  
XX  
XX New fusion protein, useful for killing diseased cells e.g. cancer cells,  
PT comprises fused first and second units, such that the protein assembles  
PT into complex which optimizes activity of multimeric form of second unit.  
XX  
XX Example 2; Fig 4B; 88pp; English.  
XX  
CC The invention relates to a fusion protein (I) comprising a first member  
CC fused to a second member, where the first and second members are chosen  
CC such that (I) assembles into a complex having a number of subunits which  
CC optimizes activity of the multimeric form of the second member. (I) is  
CC useful for selectively killing an aberrant or diseased cell which  
CC expresses a target antigen on its surface, e.g., a cancer cell expressing  
CC a cell surface antigen. (I) is also useful for detecting in vitro or in  
CC vivo the presence of target antigen in a sample, e.g., for diagnosing a  
CC disease, by contacting a sample or a control sample that allows  
CC interaction of (I) which is labeled, and detecting formation of a  
CC complex. (I) is also useful for selectively directing (e.g., localizing)  
CC the second unit of (I) to the vicinity of an undesirable cell. The  
CC present sequence represents a nucleotide sequence of a heavy chain  
CC construct 431A that was generated using the heavy chain nucleotide  
CC sequence from a humanised Mab against carcinoembryonic antigen (431)  
XX  
SQ Sequence 7328 BP; 1757 A; 1992 C; 1891 G; 1684 T; 0 U; 4 Other;  
  
Query Match 10.2%; Score 172.4; DB 4; Length 7328;  
Best Local Similarity 52.1%; Pred. No. 1.6e-42;  
Matches 492; Conservative 0; Mismatches 431; Indels 21; Gaps 4;  
  
QY 737 TGGAACTTGAAGAAAGACGAGTACCTCGACATCGGAATCAGAACGATCAGCTGGGACG 796  
DB 2825 TGGGGCCCTGTCTGACTTCTACACACTCCCTGTGGGATCCGCACTGTGGCTGTACCA 2884  
QY 797 AGAAGAGGCTTATCTGAACGGGAAACCTCTTTTGAAGGGCTTTGGAAGACGACGAG 856  
DB 2885 AGAGCCAGTTCTCATCAATGGGAAACCTTTTATTCACGCTGTCAACAAAGCATGAG 2944  
QY 857 AATCCCGCTTCTGGGGGACGGGACCTTTTATCCATGATGATTAAGACCTCAACCTTC 916  
DB 2945 ATGGGACATCCGAGGGAAGGGCTTCGACTGGCCGCTGTGTGGAAGGACTTCAACCTGC 3004  
QY 917 TGAAGTGAATCAACGGCAATTTCTTCAGAGCCTCTCACTATCTTACAGTGAAGAGTGGC 976  
DB 3005 TTCGCTGGCTTGTGTGCAACGCTTTCGATACAGCCACTACCCCTATGCAAGAGAAATGA 3064  
QY 977 TGGATCTTGGCCGACAGACTCGGAATCTTGTGATGAGCAAGCCCGGACGTTGGTATCA 1036  
DB 3065 TGCAGATGTGTACCGCTATGGGATGTGTGATCATGATGATGTCCCGGCTGGGCTGG 3124  
QY 1037 CAAGGTACCA---CTACAATCCCGAGACTCAGAAAGATAGCAAGAACATTAAGAAGA 1093  
DB 3125 CGCTGCCGAGTCTTCAACAACGTTTCTGTGATCACCACATGACGAGTGAAGAAAG 3184  
QY 1094 TGATGCAGACACAAAGAACATCCCAAGTGTATCATGTGAGAGTGTGGGAAAGAACAG 1153  
DB 3185 TGTGCGTAGGGGACAAAGAACACCCCGCGGTGTGATGTGTGTGTGGCCAAAGAGCTTG 3244  
QY 1154 AGTCCAAACCATCCAGACGGGAGGGTTTCTTCAAGCCCTTTATGAGACTGCCAATGAA 1213  
DB 3245 CGTCCACACTAGAAATCTGTGGCTACTACTTGAAGATGTGATCGCTCACACCAATCTCT 3304  
QY 1214 TGGATCGAAGACGCGCCGTTGTATGATGATGATGATGATGATGATGATGATGATGAT 1273  
DB 3305 TGGACCCCTCCCGGCTGTGACCTTGTGAGCAACTTAATATGACAG----- 3352

QY 1274 GAGACGTGGCGCTGAAGTACTTTCGACATCGTCTGTGTGAGACAGGTACTACGGCTGGTACA 1333  
DB 3353 CAGACAAAGGGGGCTCCGTATGTGATGTGATCTGTGTAACAGCTACTACTCTTGTATC 3412  
QY 1334 TCTATCAGGGAAGATAGAAAGAGACTTCAAGCTCTGAAAAAGCATAGAGAGCTCT 1393  
DB 3413 ACCACTACGGGACCTGAGTGTATTCAGCTGCAGCTGGCCACCACTTTGAGAACTGGT 3472  
QY 1394 ATGCAAGGCACAGAAAGCCCATCTTTGTACAGAAATTCGGTGGGACCGCATAGCTGCA 1453  
DB 3473 ATAAAGATATCAGAAAGCCCATTAATTCAGAGCGAGTATGAGCAGAAAGATTGACGGGT 3532  
QY 1454 TCCACTACGATCCACCTCAATGTCTTCCGAGAGTACCAAGCAGAGCTCGTTGAAAG- 1512  
DB 3533 TTCACCAAGATCCACCTCTGATGTTCACCTGAAGAGTACCAAGAAAGTCTGTAGAGAGT 3592  
QY 1513 ACATCAGGCTCTTTTGAAGAA---AGACTACATCATCGGAACACACGTTGGGCT 1567  
DB 3593 ACCATCTGGCTGTGATCAAAAACGACAGAAATATGTGTTGAGAGCTCATTTGGAAT 3652  
QY 1568 TTGCAGATTTTAAGACTCTCAGAAATGTGAGAAAGCCCATTTCAACCAAGGGTGT 1627  
DB 3653 TTGCCGATTTCAATGATGAAACAGTCAACGACGAGAGTGTCTGGGAAATTAAGGGGATCT 3712  
QY 1628 TCACAGAGACAGAACCCAACTCGTTGCTCATGTAAGTGA 1671  
DB 3713 TCACTCGGACAGACAAACAAAGTGCAGCGTTCCTTTGCGA 3756

RESULT 15

AAQ58896  
ID AAQ58896 standard; DNA; 3314 BP.

AC AAQ58896;  
XX  
DT 25-MAR-2003 (revised)  
DT 26-OCT-1994 (first entry)

XX Humanised anti-CEA sFv fragment-human beta-glucuronidase fusion gene.

XX Carcinoembryonic antigen; single chain variable region; sFv fragment;  
KW fusion gene; cancer treatment; targeted drug delivery; tumour; ss.

OS Homo sapiens.

XX Key Location/Qualifiers  
FH 1.189  
FT exon

FT /tag= a  
FT /codon\_start= 145  
FT /note= "encodes amino acids -19 to -5"

FT /tag= b  
FT /number= 2  
FT /note= "encodes amino acids -4 to 240"

FT /tag= c  
FT /number= 3  
FT /note= "encodes amino acids 241 to 257"

FT /tag= d  
FT /number= 4  
FT /note= "encodes amino acids 258 to 890 and includes the  
TAA termination codon"

PN EP590530-A2.

PD 06-APR-1994.

PF 24-SEP-1993; 93EP-00115418.

PR 02-OCT-1992; 92DE-04233152.

PA (BEHW ) BEHRINGWERKE AG.

```
XX PI Gehrmann M, Seemann G, Bosslet K, Czech J;
XX DR MPI; 1994-111012/14.
XX DR P-PSDB; AAR50092.
XX PT New fusion protein contg. enzyme for prodruq activation - coupled to
XX PT antigen binding component, esp. sfv antibody fragment, partic. for
XX PT treatment of tumours.
XX PS Claim 16; Page 12-15; 35pp; German.
XX CC The sequence A058896 codes for a humanised sfv-fragment against CEA and
XX CC a human beta-glucuronidase. The resulting fusion protein is useful for
XX CC targeting beta-glucuronidase to cancer cells expressing CEA, where the
XX CC enzyme is able to convert a prodruq into its active form. Any fusion
XX CC protein not bound to tumour can be removed by internalisation via the
XX CC mannose-6-phosphate and galactose receptors. (Updated on 25-MAR-2003 to
XX CC correct PN field.)
XX SQ Sequence 3314 BP; 789 A; 924 C; 890 G; 711 T; 0 U; 0 Other;

Query Match          10.1%; Score 170.8; DB 2; Length 3314;
Best Local Similarity 52.0%; Pred. No. 3.3e-42;
Matches 491; Conservative 0; Mismatches 432; Indels 21; Gaps 4;

QY 737 TGGAACTTGAAMAAAGACGAGTACCTCTGACATCGGAATCAGAACGATCAGCTGGAGC 796
DB 2146 TGGGGCTGTGTGTGACTTACACACTCCCTGTGGGATCCGACTGTGCTGCACCA 2205

QY 797 AGAAGAGGCTCTATCTGAACGGGAACTGTCTTTTGAAGGGCTTTGAAAGCAGAG 856
DB 2206 AGAGCCAGTTCCTCATCATGGAACCTTCTATTCCACGGGTCAACAAGCATGAG 2265

QY 857 AATTCCTCCGTTCTGGGGCAGGGACCTTTATCATTTGATGATATAAAGACTTCAACCTTC 916
DB 2266 ATGCGGACATCCGAGGGAAGGCTTCGACTGCGCGCTGTGTGAAGACTTCAACCTGC 2325

QY 917 TGAAGTGATCAAGCGGAATCTTTCAAGACCTCTCACTATCCTTACAGTGAAGTGGC 976
DB 2326 TTGCTGTGCTGTGTGCCAAGCGCTTCCGTACCAAGCACTACCCCTATGCAGAGAACTGA 2385

QY 977 TGGATCTTGCCGACAGACTCGGAATCTTGTGATAGACGAAGCCCGACGTTGGTATCA 1036
DB 2386 TGCAGATGTGTACCGCTATGGATGTGTGTCATGATGAGTGTCCCGCGCTGGCCTGG 2445

QY 1037 CAAGGTACCA---CTACAATCCCGAGACTCAGAGATAGCAGAGACAACATAGAGAA 1093
DB 2446 CGCTGCCGCACTTCTTCAACAACGTTTCTCTGCAATCACCACATGCAAGTGTGAAGAAG 2505

QY 1094 TGAATGACAGACACAAGAACCATCCAGTGTGATCATGTGAGTGTGGGAAACGAACCAAG 1153
DB 2506 TGTGCGGTAGGGACAAGAACCAACCCCGCGGTGTGATGTGTGTGTGCGCAACGAGCCTG 2565

QY 1154 AGTCCCAACCATCCAGACGCGGAGGTTTCTCAAAAGCCCTTTATGAGACTGCCAATGAAA 1213
DB 2566 CGTCCCACTTGAATCTGTGCTGCTACTACTTGAAGATGTGTGATGCTCACACCAATCCT 2625

QY 1214 TGGATCGAACACGCCCCGTTGTCTGTGATGATGATGAGCGCACGAGAGAACAA 1273
DB 2626 TGGACCCCTCCCGGCTGTGACCTTTGTGAGCAACTTAATCTATGCAG----- 2673

QY 1274 GAGACGTGGCGCTGAAGTACTTTCACATCGTCTGTGTGAACAGGTACTACGGCTGGTACA 1333
DB 2674 CAGACAAAGGGGCTCCGTATGTGTGATGTGATCTGTGTAACAGCTACTACTCTTGGTATC 2733

QY 1334 TCTATCAGGGAAGGATAGAAAGAACTTCAAGCTCTGAAAAAAGACATAGAGAGCTCT 1393
DB 2734 ACGACTACGGGCACTGAGTGTGATTCAGCTGCAGCTGGCCACCAAGTTTGAGAACTGCT 2793

QY 1394 ATGCAAGGCACAGAAAGCCCATCTTTGTCTACAGAAATTCGGTGGGACGCGATAGCTGCA 1453
DB 2794 ATTAAGAGTATCAGAAAGCCCATTAATTCAGAGCGAGTATGAGCAGAAACGATTGCAGGGT 2853
```

```
QY 1454 TCCACTACGATCCACCTCAAAATGTCTTCCGAGAGATACCAAGCAGAGCTCGTTGAAAAG- 1512
DB 2854 TTCACCAGGATCCACCTCTGATGTCTACTGAGAGTACCAAGAAAAGTCTGCTAGAGCAGT 2913

QY 1513 ACGATCAGGCTCCTTTTGA AAAA-----AGACTACATCATCGGAACACACGTTGGGCCCT 1567
DB .2914 ACCATCTGGGTCTGGATCAAAAACGAGAAAATATGTGTGAGAGACTCATTTGGAATT 2973

QY 1568 TTGCAGATTTTAAGACTCCTCAGAAATGTGAGAGAACCCATTCTCAACCAAGAGGTGTTT 1627
DB 2974 TTGCCGATTTCACTGACTGAACAGTCAACCGACGAGAGTGTGCGGATTA AAAAGGGGATCT 3033

QY 1628 TCACAAGAGACACAAACCCAACTCGTTGCTCATGTACTGAGA 1671
DB 3034 TCACTCGGACAGACAAACCAAAAAGTGACGCGTTCTTTGCCGA 3077
```

Search completed: January 24, 2005, 03:56:03  
Job time : 866 secs

***This Page Blank (uspto)***

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 24, 2005, 03:20:50 ; Search time 156 Seconds  
(without alignments)  
7695.657 Million cell updates/sec

Title: US-09-936-759-14  
Perfect score: 1689  
Sequence: 1 atggttaagaccgcaacga...gaagactgtgagtgagtc 1689

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 824507 seqs, 355394441 residues

Total number of hits satisfying chosen parameters: 1649014

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA:  
1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq:\*  
4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq:\*  
5: /cgn2\_6/ptodata/1/ina/PCTUS\_COMB.seq:\*  
6: /cgn2\_6/ptodata/1/ina/backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1688	99.9	1689	4 US-09-270-957-14	Sequence 14, Appl
2	172.4	10.2	1956	4 US-09-715-858-1	Sequence 1, Appl
3	172.4	10.2	2191	3 US-09-039-555B-12	Sequence 12, Appl
4	159.6	9.4	1947	4 US-09-715-858-3	Sequence 3, Appl
5	141.4	8.4	1887	3 US-09-149-727-7	Sequence 7, Appl
6	130.4	7.7	1888	4 US-09-270-957-27	Sequence 27, Appl
7	112	6.6	14683	2 US-08-819-866-1	Sequence 1, Appl
8	112	6.6	14683	2 US-09-023-715-1	Sequence 1, Appl
9	112	6.6	14683	4 US-09-343-485A-1	Sequence 1, Appl
10	110.8	6.6	3035	2 US-08-723-624-18	Sequence 18, Appl
11	110.8	6.6	3169	3 US-08-630-820-5	Sequence 5, Appl
12	110.8	6.6	3169	4 US-09-273-453-5	Sequence 5, Appl
13	110.8	6.6	3824	2 US-08-723-624-19	Sequence 19, Appl
14	110.8	6.6	4652	4 US-09-893-525-36	Sequence 36, Appl
15	110.8	6.6	5390	4 US-09-893-525-41	Sequence 41, Appl
16	110.8	6.6	5418	4 US-09-893-525-38	Sequence 38, Appl
17	110.8	6.6	11978	4 US-09-792-568-8	Sequence 8, Appl
18	110.8	6.6	12438	4 US-09-792-568-9	Sequence 9, Appl
19	110.8	6.6	18986	2 US-08-819-866-2	Sequence 2, Appl
20	110.8	6.6	18986	2 US-09-023-715-2	Sequence 2, Appl
21	110.8	6.6	18986	4 US-09-343-485A-2	Sequence 2, Appl
22	110.8	6.6	32798	4 US-09-604-694B-1	Sequence 1, Appl
23	109	6.5	1806	4 US-09-270-957-7	Sequence 7, Appl
24	109	6.5	1854	3 US-09-149-727-3	Sequence 3, Appl
25	109	6.5	2100	4 US-09-270-957-1	Sequence 1, Appl
26	109	6.5	6029	3 US-09-149-727-1	Sequence 1, Appl
27	104.4	6.2	8012	3 US-09-182-117-1	Sequence 1, Appl

28	104.4	6.2	8012	4 US-09-434-039A-1	Sequence 1, Appl
29	104.4	6.2	8418	3 US-09-182-117-5	Sequence 5, Appl
30	104.4	6.2	8418	4 US-09-434-039A-5	Sequence 5, Appl
31	104.4	6.2	8798	3 US-09-182-117-4	Sequence 4, Appl
32	104.4	6.2	8798	4 US-09-434-039A-4	Sequence 4, Appl
33	98	5.8	2141	4 US-09-445-283C-27	Sequence 27, Appl
34	98	5.8	2633	1 US-08-452-267-2	Sequence 2, Appl
35	98	5.8	2633	3 US-09-123-644-2	Sequence 2, Appl
36	98	5.8	4544	4 US-09-488-270A-1	Sequence 1, Appl
37	98	5.8	4947	4 US-09-118-276-21	Sequence 21, Appl
38	98	5.8	5534	1 US-08-452-267-3	Sequence 3, Appl
39	98	5.8	5534	3 US-09-123-644-3	Sequence 3, Appl
40	98	5.8	5560	3 US-08-817-188-5	Sequence 5, Appl
41	98	5.8	5642	1 US-08-318-772A-2	Sequence 2, Appl
42	98	5.8	5897	3 US-09-097-319A-26	Sequence 26, Appl
43	98	5.8	5897	4 US-09-643-971-26	Sequence 26, Appl
44	98	5.8	6898	3 US-09-097-319A-27	Sequence 27, Appl
45	98	5.8	6898	4 US-09-643-971-27	Sequence 27, Appl

ALIGNMENTS

RESULT 1  
US-09-270-957-14  
; Sequence 14, Application US/09270957  
; Patent No. 6641996  
; GENERAL INFORMATION:  
; APPLICANT: Richard A. Jefferson and Jorge E. Mayer  
; TITLE OF INVENTION: MICROBIAL - GLUCURONIDASE GENES, GENE  
; TITLE OF INVENTION: PRODUCTS, AND USES THEREOF  
; FILE REFERENCE: 190106.405C1  
; CURRENT APPLICATION NUMBER: US/09/270,957  
; CURRENT FILING DATE: 1999-03-17  
; NUMBER OF SEQ ID NOS: 112  
; SOFTWARE: FASTSEQ for windows Version 4.0  
; SEQ ID NO 14  
; LENGTH: 1689  
; TYPE: DNA  
; ORGANISM: Thermotoga maritima  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)..(1689)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-270-957-14

Query Match	99.9%	Score 1688;	DB 4;	Length 1689;
Best Local Similarity	100.0%	Pred. No. 0;		
Matches 1689;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	ATGTTAGACCCGCAACGAAACAGAGATTATTTCTTATCTTGAATGAGTTTGAAT	60	
DB	1	ATGTTAGACCCGCAACGAAACAGAGATTATTTCTTATCTTGAATGAGTTTGAAT	60	
QY	61	CTTGAAGTACCAACCAAGACAGACCAATCGCCGTTCTCTGGAAGCTGGAATGACAGTAC	120	
DB	61	CTTGAAGTACCAACCAAGACAGACCAATCGCCGTTCTCTGGAAGCTGGAATGACAGTAC	120	
QY	121	CAGGATCTGTGCTACGAGAGGACCCTTACCTTACCAAAACCACTTACGTTCCGAAG	180	
DB	121	CAGGATCTGTGCTACGAGAGGACCCTTACCTTACCAAAACCACTTACGTTCCGAAG	180	
QY	181	NAACCTTACAAACACATCAGACTTCTTGTGCGGTGAACACGAGCTCGAGGTC	240	
DB	181	NAACCTTACAAACACATCAGACTTCTTGTGCGGTGAACACGAGCTCGAGGTC	240	
QY	241	TTCTCTCAACGAGAGAAAGTGGAGAGAAATCATTGAATACCTTCCCTTCAAGTAGAT	300	
DB	241	TTCTCTCAACGAGAGAAAGTGGAGAGAAATCATTGAATACCTTCCCTTCAAGTAGAT	300	
QY	301	GTGACGGGGGAAAGTGAATCCGAGAGAGAACTCAGGGGTGTTGTGAGAACAGATTG	360	
DB	301	GTGACGGGGGAAAGTGAATCCGAGAGAGAACTCAGGGGTGTTGTGAGAACAGATTG	360	

```

QY 361 AAAGTGGAGAGATTTCCTCGAAGGTTCCAGACAGCGGCACTCACACCGTGGATTTTTT 420
DB 361 AAAGTGGAGAGATTTCCTCGAAGGTTCCAGACAGCGGCACTCACACCGTGGATTTTTT 420
QY 421 GGAAGTTTCCACCTGCAAACTTCGACTTCTCCCTACCGTGAATCATAGGCGCTGT 480
DB 421 GGAAGTTTCCACCTGCAAACTTCGACTTCTCCCTACCGTGAATCATAGGCGCTGT 480
QY 481 CTGATAGAGTTCAAGACCCAGAGATCTCGACATCTGGGTGACACAGAGTGAAGTCT 540
DB 481 CTGATAGAGTTCAAGACCCAGAGATCTCGACATCTGGGTGACACAGAGTGAAGTCT 540
QY 541 GAACCGAGAGAAACTTGAAGAAAGTGAAGATAGAAAGTCTCAGAAAGAGCGGTG 600
DB 541 GAACCGAGAGAAACTTGAAGAAAGTGAAGATAGAAAGTCTCAGAAAGAGCGGTG 600
QY 601 GGACAGAGAGATGACGATCAAACTTGAGAGAGAGAGAAAGATTAGAACATCCACAGA 660
DB 601 GGACAGAGAGATGACGATCAAACTTGAGAGAGAGAGAAAGATTAGAACATCCACAGA 660
QY 661 TTGCTGAAGGGGAGTTTCATCCTCGAAAAAGCCAGGTTCTGAGACCTCGAAGATCCATAT 720
DB 661 TTGCTGAAGGGGAGTTTCATCCTCGAAAAAGCCAGGTTCTGAGACCTCGAAGATCCATAT 720
QY 721 CTTTATCCTCTCAAGGTGGAATTGAAAAAGACAGTACACTCTGACATCGGAATCAGA 780
DB 721 CTTTATCCTCTCAAGGTGGAATTGAAAAAGACAGTACACTCTGACATCGGAATCAGA 780
QY 781 ACGATCAGCTGGGACGAGAGAGGCTCTATCTGAACGGGAAACCTGTCTTTTGAAGGCG 840
DB 781 ACGATCAGCTGGGACGAGAGAGGCTCTATCTGAACGGGAAACCTGTCTTTTGAAGGCG 840
QY 841 TTTGGAAGCAAGGAATCCCCGTTCTGGGGCAGGGCACTTTTATCATTTGATGATA 900
DB 841 TTTGGAAGCAAGGAATCCCCGTTCTGGGGCAGGGCACTTTTATCATTTGATGATA 900
QY 901 AAAGACTTCAACCTTCTGAAGTGATCAACGGAATTTTTCAGACCTTCACTATCT 960
DB 901 AAAGACTTCAACCTTCTGAAGTGATCAACGGAATTTTTCAGACCTTCACTATCT 960
QY 961 TACAGTGAAGAGTGGCTGATCTTCCGACAGACTCGGAATCTTGTGATAGCAGAGCC 1020
DB 961 TACAGTGAAGAGTGGCTGATCTTCCGACAGACTCGGAATCTTGTGATAGCAGAGCC 1020
QY 1021 CCGCAGCTGGTATCACAGAGTACCACATCAATCCCGAGACTCAGAAAGATAGCAGAGAC 1080
DB 1021 CCGCAGCTGGTATCACAGAGTACCACATCAATCCCGAGACTCAGAAAGATAGCAGAGAC 1080
QY 1081 AACATAAGAGAATGATCGACAGACACAGAACCATCCCACTGTGATCATGTGAGTGTG 1140
DB 1081 AACATAAGAGAATGATCGACAGACACAGAACCATCCCACTGTGATCATGTGAGTGTG 1140
QY 1141 GCGAAGCAACAGAGTCCAAACCATCCAGACGCGGAGGGTTTCTTCAAGCCCTTATAG 1200
DB 1141 GCGAAGCAACAGAGTCCAAACCATCCAGACGCGGAGGGTTTCTTCAAGCCCTTATAG 1200
QY 1201 ACTGCCAATGAATGATCGAACAACGCGCGTTGTCAATGATGAGCATGATGAGCAGCA 1260
DB 1201 ACTGCCAATGAATGATCGAACAACGCGCGTTGTCAATGATGAGCATGATGAGCAGCA 1260
QY 1261 GACGAGAGAAACAAGAGACGCGCGCTGAAGTACTTGCACATCGTCTGTGAACAGGTAC 1320
DB 1261 GACGAGAGAAACAAGAGACGCGCGCTGAAGTACTTGCACATCGTCTGTGAACAGGTAC 1320
QY 1321 TACGCTGCTATCATCTATCAAGGAGAGATAGAGAGACTTCAAGCTCTGAAAAAGAC 1380
DB 1321 TACGCTGCTATCATCTATCAAGGAGAGATAGAGAGACTTCAAGCTCTGAAAAAGAC 1380
QY 1381 ATAGAAGAGCTCTATGCAAGGACAGAAAGCCCATCTTTGTCAAGAAATTCGGTCCGAC 1440
DB 1381 ATAGAAGAGCTCTATGCAAGGACAGAAAGCCCATCTTTGTCAAGAAATTCGGTCCGAC 1440

```

```

QY 1441 GCGATAGCTGGCATCCATCAGATCCACTCAAAATGTTCTCCGAAGAGTACCAAGCAGAG 1500
DB 1441 GCGATAGCTGGCATCCATCAGATCCACTCAAAATGTTCTCCGAAGAGTACCAAGCAGAG 1500
QY 1501 CTCGTTGAAAAAGACGATCAGGCTCTTTGAAAAAGACTACATCATCGGAACACAGCTG 1560
DB 1501 CTCGTTGAAAAAGACGATCAGGCTCTTTGAAAAAGACTACATCATCGGAACACAGCTG 1560
QY 1561 TGGGCTTTGACAGATTTTAAAGCTCTCAGAAATGTGAGAAAGACCAATCTCAACACAG 1620
DB 1561 TGGGCTTTGACAGATTTTAAAGCTCTCAGAAATGTGAGAAAGACCAATCTCAACACAG 1620
QY 1621 GGTGTTTCAACAAGACAGACCAACCAACTCGTTGCTCATGTACTGAGAGAGCTGTG 1680
DB 1621 GGTGTTTCAACAAGACAGACCAACCAACTCGTTGCTCATGTACTGAGAGAGCTGTG 1680
QY 1681 AGTGAGTT 1689
DB 1681 AGTGAGTT 1689

```

## RESULT 2

US-09-715-858-1

; Sequence 1, Application US/09715858

; Patent No. 6582692

; GENERAL INFORMATION:

; APPLICANT: Podsakoff, Gregory

; APPLICANT: Watson, Gordon

; APPLICANT: Couto, Linda B.

; APPLICANT: Yang, Bin

; TITLE OF INVENTION: RECOMBINANT ADENO-ASSOCIATED VIRUS VIRIONS FOR THE

; FILE REFERENCE: 0800-0021

; CURRENT APPLICATION NUMBER: US/09/715, 858

; CURRENT FILING DATE: 2000-11-14

; NUMBER OF SEQ ID NOS: 16

; SOFTWARE: Patentin Ver. 2.0

; SEQ ID NO 1

; LENGTH: 1956

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (1)..(1956)

; US-09-715-858-1

Query Match 10.2%; Score 172.4; DB 4; Length 1956;

Best Local Similarity 52.1%; Pred. No. 2.1e-43;

Matches 492; Conservative 0; Mismatches 431; Indels 21; Gaps 4;

```

QY 737 TGAAGCTTGAAGAAAGACGAGTACCTGTGACATCGGAATCAGAGATCAGTGGAGC 796
DB 737 TGGGCTGTGTGCTGACTTCTACACACTCCCTGTGGGATCCGCACTGTGCTGACCA 997
QY 797 AGAAGAGCTCTATCTGAACGCGAAACCTGTCTTTTGAAGGCTTTGAAAGCAGAG 856
DB 797 AGAGCAGTCTCATCAATGGAACCTTTCTATTTCCACGCTGTCAACAAGCATGAGG 1057
QY 857 AATTCCCGTTCTGGGCGAGGCACTTTTATTCATGTGATGATGATAAAGACTTCAACTTC 916
DB 857 ATGCGACATCCGAGGGAAGGCTTCGACTGCGCGCTGCTGTGAAGAGACTTCAACTTC 1117
QY 917 TGAATGATCAACGCGAATCTTTCAGAGCTCTCACTATCTTACAGTGAAGAGTGC 976
DB 917 TTGCTGCTGTGTGCTCAACGCTTTCGTACACGCACTACCCCTATGACAGAGAGTGA 1177
QY 977 TGAATCTTCCGACAGACTCGAATCTTGTGATAGAGAGAGCCCGCAGTGGTATCA 1036
DB 977 TGCAGATGTGTACCGCTATGGGATGTGTGATGATGATGATGATGATGATGATGATG 1237
QY 1037 CAAGTACCA---CTACAATCCGAGACTCAGAGATAGCAGAGACACATAGAGAA 1093
DB 1037 CGCTGCCGCGAGTTCTTCAACAAGGTTTCTCTGCATCACCAATGACAGTATGAGAGAA 1297

```



[illegible]

RESULT 3  
US-09-039-555B-12  
; Sequence 12, Application US/09039555B  
; Patent No. 6033856  
; GENERAL INFORMATION:  
; APPLICANT: Koerner, Kathrin  
; APPLICANT: Mueller, Rolf  
; APPLICANT: Sadlacek, Hans-Harald  
; TITLE OF INVENTION: PROMOTER OF THE CDC25B GENE, ITS  
; TITLE OF INVENTION: PREPARATION AND USE  
; NUMBER OF SEQUENCES: 19  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 3000 K Street, N.W., Suite 500  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20007-5109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/039,555B  
; FILING DATE: 16-MAR-1998  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: DE 19710643.9  
; FILING DATE: 14-MAR-1997  
; ATTORNEY/AGENT INFORMATION:  
;

```

; NAME: Bent, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 016779/0131
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2191 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
;
US-09-039-555B-12

```

Query Match	10.2%;	Score 172.4;	DB 3;	Length 2191;
Best local Similarity	52.1%;	Pred. No. 2.3e-43;		
Matches 492; Conservative	0;	Mismatches 431;	Indels 21;	Gaps 4;

QY	737	TGGAACCTTGAAAAAGACGAGTACACTCTGACATCGGAATCAGAAACGATCAGCTGGGACG	796
Dp	964	TGGGGCCTGTGTTCTGACTTCTPACACACTCCCTGTGGGATCGCACTGTGGCTGTACCA	1023
QY	797	AGAAAGAGCCTTATCTGAAACGGGAAACCTGTCTTTTGAAGGCTTTGAAACACGAGG	856
Dp	1024	AGAGCCAGTTCCTCATCAATGGGAAACCTTTCTATTTCACGCGTGTCAACAAGCATGAGG	1083
QY	857	AATTCCTCCGTTCTGGGGCAGGGCACTTTTATCCATTGATGATAAAGACTTCAACCTTC	916
Dp	1084	ATGCGGACATCCGAGGGAGGGCTTCGACTGGCCGCTGTGGTGAAGACTTCAACCTGC	1143
QY	917	TGAAGTGATCAACGCAATCTTTCAGAGCTCTCACTATCTTACGTGTAAGAGTGGC	976
Dp	1144	TTGCTGGCTTGGTGGCCAAACGCTTTCGTTACCAACCACTACCTTATGACAGGAAAGTGA	1203
QY	977	TGGATCTTGGCCGACAGACTCGGAATCCTTGTGATAGACGAAGCCCGACGTTGTATCA	1036
Dp	1204	TGCAGATGTGTACCGCTATGGGATTTGTGTATCGATGATGTGCCGCGTGGGCTGG	1263
QY	1037	CAAGTACCA---CTACAATCCCGAGACTCAGAAAGATAGCAGAACAAATPAAAGAA	1093
Dp	1264	CGCTGCCGAGTTCTTCAACAACGTTTCTGTGCATCACCAATGACAGTATGAAAGAAAG	1323
QY	1094	TGATCGACAGACACAAGAACCATCCAGTGTGATCATGTGAGTGTGGCGAAGCAACGAG	1153
Dp	1324	TGTTGCGTAGGGACAAGAACCAACCCGCGGTGTGATGTGTCTGTGGCAACGAGCCTG	1383
QY	1154	AGTCCAACCATCCAGACGCGGAGGTTTCTTCAAAAGCCCTTATGAGACTGCCAATGAAA	1213
Dp	1384	CGTCCACCTAGAAATCTGTCTGGCTACTACTGAAAGATGTGATCGCTCACACCAAAATCCT	1443
QY	1214	TGGATCGAACACGCCCCGTTGTCTATGTGTAGCATGATGACGCAACGACGAGAGAACAA	1273
Dp	1444	TGGACCCCTCCCGGCTGTGACCTTTGTGAGCAACTTAATATGACG-----	1491
QY	1274	GAGACGTGGCGCTGAAGTACTTCGACATCGTCTGTGTAACAGCTACTACGGCTGTACA	1333
Dp	1492	CAGACAAAGGGGCTCCGTATGTGATGTGATCTGTTTGAACAGCTACTACTTGTGTATC	1511
QY	1334	TCTATCAGGGAAGATAGAAAGAGACTTCAAGCTCTGAAAAAAGACATAGAAAGCTCT	1393
Dp	1552	ACGACTACGGGCACCTGAGTTGATTACGTGCAGCTGGCCACCAGTTTGAAACTGTGT	1611
QY	1394	ATGCAAGGCACAGAAAGCCCACTTTGTGCAGAAATTCGTGGCGAGCGGATAGCTGCA	1453
Dp	1612	ATAAAGATATCAGAAAGCCCAATTATTACAGAGCAGATATGGAGCAGAAACGATTCAGGGT	1671
QY	1454	TCCACTACGATCCACTCAATATGTTCTCCGAAAGATACCAAGCAGAGCTGTTGAAAAG-	1512
Dp	1672	TTCAACCAAGATCCACCTGTGATGTTCACTGAAGAGTACCAAGAAAAGTCTGTGATGAGCAGT	1731
QY	1513	ACGATCAGGCTCTTTGAAAAA----AGACTACATCATCGGAACACACAGTGTGGGCT	1567

```

Db      1732 ACCATCTGGGTCTGATCAAAAACGAGAAAATATGTGTTGAGAGCTCATTTGGAATT 1791
QY      1568 TTGCAGATTTTAAGACTCTCAGATGTGAGAACCCATTCTCAACCAAGGTTT 1627
Db      1792 TTGCCGATTTTCATGACTGACAGTCAACCGAGAGTGTCTGGGAAATAAAAAGGGATCT 1851
QY      1628 TCACAGAGACAGAACCAACCAACTGTTGCTCATGTACTAGA 1671
Db      1852 TCACTCGGACAGACAAACCAAAAGTGCAGCGTCTTTTCCGA 1895

```

## RESULT 4

```

US-09-715-858-3
; Sequence 3, Application US/09715858
; Patent No. 6582692
; GENERAL INFORMATION:
; APPLICANT: Podsakoff, Gregory
; APPLICANT: Watson, Gordon
; APPLICANT: Couto, Linda B.
; APPLICANT: Yang, Bin
; TITLE OF INVENTION: RECOMBINANT ADENO-ASSOCIATED VIRUS VIRIONS FOR THE
; FILE REFERENCE: 0800-0021
; CURRENT APPLICATION NUMBER: US/09/715,858
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 3
; LENGTH: 1947
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1947)
; US-09-715-858-3

```

Query Match 9.4%; Score 159.6; DB 4; Length 1947;

Best Local Similarity 50.6%; Pred. No. 2.2e-39;

Matches 477; Conservative 0; Mismatches 444; Indels 21; Gaps 3;

```

QY      739 GAACCTGAAAAGAGAGTACTCTGACATCGGAATCAGAACGATCAGCTGGAGCAG 798
Db      928 GAGTCTGTGACTGACTACTACCCCTTCTATCGGATTCGAACAGTGGCTGCACAAAG 987
QY      799 AAGAGGCTCTATCTGAACGGGAAACCTGTCTTTTGAAGGCTTTGAAAGCAGAGAA 858
Db      988 AGCAAGTCTCTATTAACGGGAAACCCCTTCTATTTCCAAAGGGGTCAATTAAGCAGAGAT 1047
QY      859 TTCCCGCTTCTGGGAGGAGCACTTTTATTCATTGATGATTAAGAACTTCAACCTTCTG 918
Db      1048 TCAGATATTCGAGGAAAGGCTTGCAGTGGCCGCTGCTGTTAAAGATTCAACCTGCTC 1107
QY      919 AAGTGATCAACGCGAATCTTTCAGAGCCTCTCACTATCCTTACAGTGAAGTGGCTG 978
Db      1108 CGTTGGCTGGGCAAAATTCCTTTCGTACAGCCACTATCCCTACTCAGAGAGTACTT 1167
QY      979 GATCTTGGCAGACAGACTCGGAATCCTTGTGATAGACGAGACCCCGCAGCTTGG--TATC 1035
Db      1168 CAGCTCTGTGACCGATACGGGATTTGTGTCATGATGATGTCCTCCGGTGTGGCATTTGTG 1227
QY      1036 ACAAGGTACCACTACATCCGAGACTCAGAAAGATAGACAGAAAGACAAATTAAGAAGATG 1095
Db      1228 CTACCTCAGAGTTTGGCAACGAGTCACTTGGCACCACCTAGAGGTGATGAGAGACTG 1287
QY      1096 ATGCACAGACAAAGAACATCCAGTGTGATCAATGTGAGTGTGGGAAAGCAAGACAGAG 1155
Db      1288 GTTCGCGGGGACAAAAATCAACCTCGGTGTGTGATGTGTGTGTGTGTGTGTGTGTGTGT 1347
QY      1156 TCCAACCATTCAGACGCGGAGGTTTCTTCAAGCCCTTTATGAGACTGCCAATGAATG 1215
Db      1348 TCTGCTCTGAAACCCCGCATATTAATTAAAGCGCTGATCAACCAACCAAAAGCCCTG 1407

```

```

QY      1216 GATGAAACACGCCCCCTTGTCTATGTGTGACATGATGAGCCGACAGACGAGAAACAAGA 1275
Db      1408 GACCTCACCCGTCCTCCGTAACCTTTGTAGCA-----ACGCCAATATATGATGCA 1455
QY      1276 GACGTGGGCTGAAGTACTTGCACATCTGTGTGTGAACAGGTACTACGGCTGTATC 1335
Db      1456 GACCTGGGGCCCCCGTACGTGATGTATCTGTGTAAACAGCTACTTTTCTTGTGTATCAT 1515
QY      1336 TATCAGGGAAGATAGAAAGAGACTTCAAGCTCTGAAAAAGACATAGAAAGCTCTAT 1395
Db      1516 GACTATGGGCATTTGGAGGTGATTCAAGCCACAGCTGAATAGCCAGTTTGAAGACTGTAT 1575
QY      1396 GCAAGGACACAGAAAGCCCATCTTTGTCAAGATTCGCTCGGACCGCATAGCTGGCATC 1455
Db      1576 AAGACGATCAGAAAGCCGATTATCCAGAGCGAGTATGAGACAGACCAATCCAGGATC 1635
QY      1456 CACTACGATCCACCTCAATATGTTCGGAAGAGTACCAAGACAGCTCGTTGAAAAAGACG 1515
Db      1636 CACGAGAACCCGCTCGCATGTTCAAGTGAAGATCAAGAGGCTGTCTGAGAAATTAC 1695
QY      1516 ATCAGGCTCTTTGAAA-----AAAGACTACATCGGAACACACGTTGGCCTTT 1569
Db      1696 CATTCAGTCTGTGATCAGAAACGTAAGATACGTGTCGAGAGCTCATCTGAAATTTT 1755
QY      1570 GCAGATTTTAAGACTCTCAGATGTGAGAAAGACCATTTCTAACCAAGAGGTTTTC 1629
Db      1756 GCCGACTTCATGACGAACAGTCAACCGCTGAGAGTAAATCGGAACAAGAGGGATCTT 1815
QY      1630 ACAAGAGACAGAACCCAACTCGTTGCTCATGTACTAGA 1671
Db      1816 ACTGCCAGAGACAGCCCAAACTTCGGCCTTATTTTGGCA 1857

```

## RESULT 5

US-09-149-727-7

; Sequence 7, Application US/09149727

; Patent No. 6391547

; GENERAL INFORMATION:

; APPLICANT: Jefferson, Richard A.

; APPLICANT: Kilian, Andrzej

; APPLICANT: Keese, Paul Konrad

; TITLE OF INVENTION: MICROBIAL BETA-GLUCURONIDASE GENES, GENE PRODUCTS AND

; FILE REFERENCE: 190106.405

; CURRENT APPLICATION NUMBER: US/09/149,727

; EARLIER FILING DATE: 1998-09-08

; EARLIER APPLICATION NUMBER: US 60/058,263

; EARLIER FILING DATE: 1997-09-09

; NUMBER OF SEQ ID NOS: 71

; SOFTWARE: Patentin Ver. 2.0

; SEQ ID NO 7

; LENGTH: 1887

; TYPE: DNA

; ORGANISM: Bacillus sp.

; US-09-149-727-7

Query Match 8.4%; Score 141.4; DB 3; Length 1887;

Best Local Similarity 45.4%; Pred. No. 1.1e-33;

Matches 734; Conservative 0; Mismatches 822; Indels 60; Gaps 4;

```

QY      120 CCAGATCTGTGCTACGAAGAGACCTTCACTACAAACCAACCTTCTACGTTCCGAA 179
Db      208 CAAGAAATCGCAACCAATATCGGATATGCTGTGACGAACGTGAGTTCAAGGTGCCGC 267
QY      180 GNAACCTTCAAAAAACATCAGACTTTTGTGCTGGGTGAACACGACTGCGAGT 239
Db      268 CTATCTGAAGATCAGCGTATCGTCTCGCTTGGCTGCAACTCAAAAGCAATTTGT 327
QY      240 CTTCCTCAACGAGAGAAAGTGGAGAGATCATTGAATACCTTCCCTTGAAGTAGA 299
Db      328 CTATGTCAATGTGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 387
QY      300 TGTGACGGGGAAGTGAATCCGAGAGAAAGAACTCAAGGTGTTGTTGAAGACAGATT 359

```

```
Db      388 AATCAACAACCTCGCTGCGTGATGCGATGATCGCGTCAACCGTCCGCGTGACAAACATCCT 447
Qy      360 GAAAGTGGAGATTTCCTCGAAGGTTCCAGACAGCGCACTCAACCGTGGATTTT 419
Db      448 CGACGATAGCAACCTCCCGTGGGCTGTACAGCGAGCGCCACGAAGAGGGCCTCGGAAA 507
Qy      420 TGGAAAGTTTCAACCTGCAAACTTCACTTCCCTACGCTGGAATCATAGGCTGT 479
Db      508 AGTCATTGTAACAAGCCGAACCTTCACTTCACTATGACAGCCCTGCACCGCCGGT 567
Qy      480 TCTGATAGATTGACAGACCAACGAGGATCTGCATCTGGGTGACACGAGTGAATC 539
Db      568 GAAATCTACACGACCCCGTTAGCTACGTCGAGGACATCTCGGTGTGACCGCACTTCAA 627
Qy      540 TGAACCGGAGAGAACTTGAAGAGTGAAGTAGATAGAGTCTCAGAGAACGGT 599
Db      628 TGGCCCAACCGGACTGTGACCTATACGCTGACTTTCAGGCAAGCCGAGACCGTGAA 687
Qy      600 GGGACAGAGATGACGATCAAACTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 659
Db      688 AGTGTCCGTGCTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 747
Qy      660 ATTCTCGAAGGGAGTTCATCTCTGCAAAACGCGAGCTTTCGACCTCGAAGATCCATA 719
Db      748 CGTGGAGATTCCGAATGTCTATCTCTGGGAAACCACTGAACACGTATCTTACCAATCAA 807
Qy      720 TCTTTATCTCTCAAGGTGGAATTTGAAAAAGACAGTACATCTGGACATCGGAATCAG 779
Db      808 AGTGAAGTGGTGAACGACGAGTACCATCGATGTCTATGAAGAGCCGTTCCGCGTCGG 867
Qy      780 AACGATCAGCTGGGACGAGAGAGGCTCTATCTGAACGGGAACTGCTTTTGAAGGG 839
Db      868 GACCGTGGAAGTCAACGACGCGCAAGTTCCTCATCAACAAACCGTTCTACTTCAAGGG 927
Qy      840 CTTTGAAGACGAGGAATTCCTGCTGCGGAGGGGACCTTTTATCCATTGATGAT 899
Db      928 CTTTGGCAACATGAGGACATCTCTATCAACGCGCGTGGCTTTAAAGAGCGACATGT 987
Qy      900 AAAAGACTTCAACCTTCTGAAGTGATCAACGCGAATCTTTCAAGGACCTCTCACTATCC 959
Db      988 GATGGATTCAATATCTCAAAATGATTCGCGCCCAACAGCTTCGGAACCGCACATATCC 1047
Qy      960 TTACAGTGAAGAGTGGCTGATCTTCCGACAGACTCGGAATCCTTGTATAGAGAAAC 1019
Db      1048 GTACTCTGAAGAGTGTATGCGTCTTGGGATCGCGAGGGTCTGCTGATCGAGAGAC 1107
Qy      1020 CCGGACGTTGGTAT----- 1034
Db      1108 TCCGGCAGTTGGCGTGACCTCAACTTCATGGCCACACGCGACTCGCGAAGCGACGGA 1167
Qy      1035 ---CACAAAGTACCACTACATCCCGAGACTCAGAAAGATAGCAGAAAGCAACATTAAGAG 1091
Db      1168 GCGCGTCAGTACTGGGAGAGAGATTGGACGTTTGAACCAATCAAGACGTTCTCCGTGA 1227
Qy      1092 AATGATCGACAGACACAAAGAACATCCAGTGTGATCATGTGAGTGTGCGAAGAAC 1151
Db      1228 ACTGCTGTCTGTGACAAAGAACATCCAGCGTGTGATGTGAGCATCGCCAAAGAGGC 1287
Qy      1152 AGAGTCCAAACATCCAGACGCGGAGGTTTCTTCAAAAGCCCTTATGAGACTGCCAATGA 1211
Db      1288 GCGCACTGAGAGAGAGGCGCGTACGACTACTTCAAGCCGTTGTGAGCTGACCAAGGA 1347
Qy      1212 AATGATCGA---ACACGCGCGTGTCTATGTCATGTCATGATGAGCATGATGAGCGACAGAG 1268
Db      1348 ACTGACCCACAGAAAGCGTCCGCTCAAGTCGTGCTGTTGTGATGCTACCCCGAGAGAC 1407
Qy      1269 AACAAAGAGACGTGGCGCTGAAGTACTTGCACATCTGTGTGAACAGGTACTACGGCTG 1328
Db      1408 GGACAAAGTCCCGAACTGA-----TTGACGTATCGCGCTCAATCGCTATAACGAGATG 1461
Qy      1329 GTACATCTATCAGGGAAGGATAGAGAG--GACTTCAAGCTCTGAAAAAGACATAGA 1385
```

```
Db      1462 GTACTTCGATGGCGGTGATCTCGAAGCGGCCAAAGTCCATCTCCGCCAGGAATTTCAAGC 1521
Qy      1386 AGAGCTTATGCAAGGCACAGAAAGCCCATCTTTGTCAAGAAATTCGGTCCGACCGCAT 1445
Db      1522 GTGGAACAAGCGTTGCCAGGAAGCCGATCATGATCATCTAGTACGGCGACAGACCGT 1581
Qy      1446 AGCTGCATTCATACGATTCACCTCAAAATGTTCTCCGAAGAGTACCAAGCAGACTCGT 1505
Db      1582 TCGCGGCTTTCACGACATTGATTCAGTATGTTTCAACGAGGAATATCAAGTCACTA 1641
Qy      1506 TGAAGGACGATCAGGCTCTTTTGAAGAACTACATCATCGAAACACACGCTGGGC 1565
Db      1642 CAGGCGAACCACTGCTGTTCGATGAGTTTGAAGAACTTCGTGGGTGAGCAAGCGTGAA 1701
Qy      1566 CTTGCAATTTTAAAGCTCTCAGAAATGTGAGAGAACCCATTTCAACCAAGGCTGT 1625
Db      1702 CTTCCGGAATTCGCGACCTTCAGAGGCGGTGATGCGCTCAAGAGAAACAAAGAGCGGT 1761
Qy      1626 TTTCACAAGAGACAGACCAACCAAACTGCTGCTCATGTACTGAAGAGACTGTGA 1681
Db      1762 GTTCACTGCTGACCGCAAGCCGAAGCTCGCGCGACGTCCTTTCGAGCGCTGA 1817
```

## RESULT 6

```
US-09-270-957-27
; Sequence 27, Application US/09270957
; Patent No. 6641996
; GENERAL INFORMATION:
; APPLICANT: Richard A. Jefferson and Jorge E. Mayer
; TITLE OF INVENTION: MICROBIAL -GLUCURONIDASE GENES, GENE
; FILE REFERENCE: 190106.405C1
; CURRENT APPLICATION NUMBER: US/09/270,957
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 27
; LENGTH: 1888
; TYPE: DNA
; ORGANISM: Bacillus sp.
US-09-270-957-27
```

```
Query Match      7.7%; Score 130.4; DB 4; Length 1888;
Best Local Similarity 45.4%; Pred. No. 3.1e-30;
Matches 734; Conservative 0; Mismatches 822; Indels 61; Gaps 5;
```

```
Qy      120 CAAGATCTGTGCTACGAGAGAGACCCCTTCACTACAAACCACTTCTACGTTCCGAA 179
Db      208 CAAGGAATCCGCAACCATATGATATGTCTGTAGCAACGTGAGTTCAACGTCGGC 267
Qy      180 GNAACTTTCACAAAAACATCAGACTTACTTTGCTGCGGTGAACACGACTCGAAGT 239
Db      268 CTATCTGAAGATCAAGCTATGTGCTCCGCTTCGGCTCTGCAACTCACAAGCAATTGT 327
Qy      240 CTTCTCAACGAGAGAGAAAGTGGAGAGATCAATTGAATACCTTCCCTTGAAGTAGA 299
Db      328 CTATGTCAATGTGAGCTGTGTGTGAGACAAAGGCGGATTCCTGCCATTGCAAGCGGA 387
Qy      300 TGTGACGGGGAAGTGAATCCGAGAGAGAACCAACTCAGGGTGTGTTGAGAACAGATT 359
Db      388 AATCAACATCTGCTGCGTGTGATGCGATGAATGCGCTCACCGTCCCGTGAACAATCCT 447
Qy      360 GAAAGTGGAGGATTTCCCTCGAAGGTTCCAGACAGCGGCACTCAACCGTGGATTTTT 419
Db      448 CGACGATAGCAACCTCCCGGTGGGCTGTACAGCGGACGAGAGAGAGAGAGAGAGAGAG 507
Qy      420 TGAAGTTTTCACCTGCAAACTTGCACCTTTCCTTCCCTTACCGTGAATCATAGGCTGT 479
Db      508 AGTCATTCTGTAACAAGCCGAACCTTCACTTCTTCAACTATGACAGGCGCTGCACCGTCCGGT 567
Qy      480 TCTGATAGATTCAAGACCAACCGGAGATCTGACATCTGGGTGACACGAGTGAATC 539
Db      568 GAAATCTACACGACCCCGTTTACGTACGTGAGAGACATCTCGGTTGTGACCACTTCAA 627
```

QY 540 TGAACCGAGAGAACTTGGAAAAGTGAAGATAGAGTCTCAGAGAAGCGGT 599  
 DB 628 TGGCCCAACCGGACTGTGACCTATACGGTGGACTTTCAAGGCAAAAGCCGAGACCGTGAA 687  
 QY 600 GGGACAGAGATGACGATCAAACTTGGAGAGAGAGAAAAAGATTAGACATCCAACAG 659  
 DB 688 AGTGTGGTGGTGGATGAGAGAGGCAAAAGTGGTCGAAGCACCAGGCGCTGAGCGGTAA 747  
 QY 660 ATTCGTGAAGGGAGTTCCTC- GAAAAGCCAGGTTTGGAGCCTCGAAGATCCAT 718  
 DB 748 CGTGAGATTCGAATGTCTATCTTGGAAACCACTGAACACGTATCTTACCAGATCA 807  
 QY 719 ATCTTATCTCTCAAGGTGAACTTGAAGAAAGAGATCACTTGACATCGGAATCA 778  
 DB 808 AAGTGAACGTGGTGAACGACGGAAGTCAATCGATGTCTATGAAGAGCCGTTCCGGTGC 867  
 QY 779 GAAAGATCAGCTGGGACGAGAGAGGCTCTATCTTGAACGGGAAACCTGTTTGAAG 838  
 DB 868 GGAACGTGAAGTCAACGACGCGCAAGTCTCTATCAACAACAAACGTTCTTCACTTCAAG 927  
 QY 839 GCTTTGAAAGCAGAGAAATCCCCGTTCTGGGGCAGGCACTTTTATCATTTGATGA 898  
 DB 928 GCTTTGGCAACATGAGAGCACTCTATCAAGCGCCGTGGCTTTAAGCAAGCGAATG 987  
 QY 899 TAAAGACTTCAACCTTCTGAAGTGAATCAACGCAATTTTTCAGACCTCTCACTATC 958  
 DB 988 TGATGATTTCAATATCTCTAATAGATGGCGCCCAACAGCTTCCGACCGCACACTATC 1047  
 QY 959 CTTACAGTGAAGAGTGGCTGATCTTGGCCGACAGACTCGGAATCTTGATAGACGAAG 1018  
 DB 1048 CGTACTCTGAAGAGTGTGATGCTTGGCGGATCGCAGAGGTCTGTCGTATCGACGGA 1107  
 QY 1019 CCCCAGCAGTTGGTAT----- 1034  
 DB 1108 CTCGGCAGTTGGCGTGCACCTCACTTCACTGAGCCACACGAGGACTCGGCGAAGCAGCG 1167  
 QY 1035 -----CACAAGGTACCACTACATCCCGAGACTCAGAGATAGCAGAGACAACATAAGAA 1090  
 DB 1168 AGCGCGTCACTACCTGGAGAGAGATTCGGACGTTTGAGACCATCAAGACGTTCTCCGTG 1227  
 QY 1091 GAATGATCGACAGACACAAGAACCATCCAGTGTGATCATGTGAGTGTGGCGAAGCAAC 1150  
 DB 1228 AACTGGTGTCTGTGACAGAAACCATCAAGCGTGTGATGTGAGATCGCCAAACGAG 1287  
 QY 1151 CAGAGTCCAAACCATCCAGAGCGGAGGTTTCTTCAAGCCCTTTATGAGACTGCCAATG 1210  
 DB 1288 CGGCGACTGAGAGAGAGGCGGCTACGAGTACTTCAAGCCGTTGTGAGCTGACCAAGG 1347  
 QY 1211 AAATGATGCA--ACACGCGCGGTTGTGATGTGAGCATGATGACGACACAGACGAGA 1267  
 DB 1348 AACTCGACCCACAGAAAGCGTCCGGTACAGATCGTGTGTGTGATGAGCTACCCCGGAGA 1407  
 QY 1268 GAACAAGAGACGTGGCGCTGAAGTACTTGCACATCGTGTGTGAAACAGTACTACGCT 1327  
 DB 1408 CGGACAAAGTCCCGCAACTGA-----TTGACGTCAATCGCGCTCAATCGCTATTAACGAT 1461  
 QY 1328 GGTACATCTATCAGGGAAGATAGAAAG--GACTTCAAGCTCTGGAAGAAAGACATAG 1384  
 DB 1462 GGTACTTGTGATGGCGGTGATCTCGAAGCGGCCAAAGTCAATCTCCGCAAGAAATTTCAG 1521  
 QY 1385 AAGAGCTTATGCAAGGCAAGAAAGCCCATCTTTGTCAAGAAATTCGTCGAGCGGA 1444  
 DB 1522 CGTGAACAAAGCGTTGCCAGGAAAGCCGATCATGATCACTGAGTACGGCGCAGACACCG 1581  
 QY 1445 TAGCTGGCATCCACTACGATCCACCTCAATGTTCTCGAAGAGATACCAAGCAGAGCTG 1504  
 DB 1582 TTGCGGGCTTTCACGACATGTATCCAGTGTGTTTACCGAAGAAATATCAAGTCACT 1641  
 QY 1505 TTGAAAAGACGATCAGGCTCTTTTGAAGAAAGACTATCATCGGAACACAGCTGTGG 1564  
 DB 1642 ACCAGGCGAACCAACGTGCTGTGATGAGTGTGAAACTTCTGTGGTGAAGCAAGCGTGA 1701

QY 1565 CCTTGGAGATTTAAGACTCCTCAGAAATGTGAGAGACCATTCTCAACCAAGGGTG 1624  
 DB 1702 ACTTCGGGACTTCGGCACTCTCAGGGCGGTGATGCGCTCCAAAGAAACAAAGAGGCG 1761  
 QY 1625 TTTTCAAGAGACAGACAAACCCAACTCGTTGCTCATGTACTGAGAAAGACTGTGA 1681  
 DB 1762 TGTCTACTGTGACCGCAAGCCGAAGCTCGCCGCGCAGCTCTTTCGCAAGCGCTGA 1818

## RESULT 7

US-08-819-866-1/c

Sequence 1, Application US/08819866

Patent No. 5830698

GENERAL INFORMATION:

APPLICANT: REEF, Mitchell E.

APPLICANT: BARNETT, Richard Spence

APPLICANT: MCLACHLAN, Karen Retta

TITLE OF INVENTION: NOVEL METHOD FOR INTEGRATING GENES AT

SPECIFIC SITES IN MAMMALIAN CELLS VIA HOMOLOGOUS

RECOMBINATION AND VECTORS FOR ACCOMPLISHING THE SAME

NUMBER OF SEQUENCES: 2

CORRESPONDENCE ADDRESS:

ADDRESSEE: BURNS, DOANE, SWECKER &amp; MATHIS, L.L.P.

STREET: P.O. Box 1404

CITY: Alexandria

STATE: Virginia

COUNTRY: United States

ZIP: 22313-1404

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/819,866

FILING DATE: 14-MAR-1997

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Teskin, Robin L.

REGISTRATION NUMBER: 35,030

REFERENCE/DOCKET NUMBER: 012712-352

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 836-6620

TELEFAX: (703) 836-2021

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 14683 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

US-08-819-866-1

Query Match

Best Local Similarity 48.1%; Pred. No. 6.7e-24;

Matches 472; Conservative 0; Mismatches 450; Indels 60; Gaps 3;

QY 757 TACACTCTGACATCGGAATGAGAACGATCAGCTGGAGCAGAAAGAGCTTATCTGAAC 816  
 DB 7348 TACCCGCTTCGCGTGGCATCGGCTCAGTGCAGTGAAGGCGAACAAGTCTGATTAAC 7289  
 QY 817 GGGAAACCTGTCTTTTGAAGGCTTTGGAAGCAGAGAAATCCCGTTTGGGGCAG 876  
 DB 7288 CACAAACCGTTCTACTTATGCTTGTGCTGTCATGAAGATGCGGACTTGGCGCAA 7229  
 QY 877 GGCACCTTTTATCCATGTGATTAAGAAAGCTTCAACCTTGAAGTGAACGCAAT 936  
 DB 7228 GGATTCGATAACGTGATGATGTGACAGACAGCATTAATGAGTATGAGGCGCAAC 7169  
 QY 937 TCTTTCAGACCTCTCACTATCCTTACAGTGAAGAGTGGCTGATCTTGGCAGACACTC 996  
 DB 7168 TCCTACCGTACTCGCATTAACCTTACGCTGAAGAGATGCTCGACTGGCAGATGAACAT 7109

QY	997	GGAAATCCTTGTGATAGACGAAGCCCC-----GCACTGTGCT	1032
DB	7108	GGCATCGTGGTGAATTGATGAACCTGCTGCTGTGGCTTTAACTCTCTTTAGGCATTGGT	7049
QY	1033	ATCACAAAGGTACCACTACAATCCCCGAGACTCAGAAGTAGCAGAAGACAAC-----	1083
DB	7048	TTCCGAAGCGGGCAACAAGCCGAAAGAACTGTACAACGAAGAGGCAAGTCAACGGGGAAACT	6989
QY	1084	-----ATAAGAAGATGATCGACACACACAAGAACCATTCCAGT	1122
DB	6988	CACCAAGCGCACTTACAGCGCGATTAAAGAGCTGATAGCGCGTGACAAAAACCAACCAAGC	6929
QY	1123	GTGATCATGTGGAGTGTGCGCAACGAACCAAGATCCAAACCATCCAGACCGCGAGGGTTTC	1182
DB	6928	GTGGTGATGTGGAGTATTGCCAACGAACCGGATACCCTCCGCAAGGTGCACCGGAATAT	6869
QY	1183	TTCAAAGCCCTTATGAGACTGCGCAATGAATGATCGAACAACGCCCCGTGTATGCTG	1242
DB	6868	TTCCGCCCACTGSCGGGAAGCAACGCGTAAACTGCACCCGACGCGTCCGATCACTGCGTC	6809
QY	1243	AGCATGATGAGCAACCAAGACGAGACGAGACAAGACGAGACGCTGGCGCTGAAGTACTTCGACATC	1302
DB	6808	AATGTATGTCTGCGGACGCTCACACCGATACATCAGCGATCT-----CTTTGATGTG	6755
QY	1303	GTCTGTGTAACAGGTACTACGCGCTGTACATCTATCAGGGAAGATAGAAGAAGACTT	1362
DB	6754	CTGTGCCCTGAACCGTTATTACGGATGGTATGTCCAAAGCGGCGATTGGAAAACGGCAGAG	6695
QY	1363	CAAGCTCTGGA AAAAGACATAGAAGAGCTCTATGCAAGGACAGAAAGCCCATCTTGTG	1422
DB	6694	AAGTACTGGA AAAAGAACTTGTGGCCTGGCAGGAAGAACTGCATCAGCCGATTATCATC	6635
QY	1423	ACAGAATTCGTGCGGACGCGAGTAGCTGGCATCCACTACGATCCACTCAATGTTCTCC	1482
DB	6634	ACCGAATACGCGGTGATACGTTAGCCGGCTGCACCTCAATGTACACCGACATGTGAGT	6575
QY	1483	GAAGAGTACCAAGCAGAGCTCGTTGAAAAAGACGATCAGGCTCTTTGAAAAAAAGACTAC	1542
DB	6574	GAAGAGTATCAGTGTGCATGGCTGGATATGTATCACCGCGCTTTGATCGGTCAGCGCC	6515
QY	1543	ATCATCGGAACACAGTGTGGGCTTTGACAGATTTTAAAGTCTCTCAGAAATGTGAGAAGA	1602
DB	6514	GTCTGTGGTGAACAAGTATGGAATTTGCGCGATTTTGCGAATCTCGCAAGGCATATTGCGC	6455
QY	1603	CCCATTTCTCAACCAAGGGTGTTCACACAAGACAGACACCAACCAACTCGTGTCTCAT	1662
DB	6454	GTTGGCGGTAAACAAGAAAGGATCTTCACTGCGGACCGCAACCGAAGTCGGCGGCTTTT	6395
QY	1663	GTAAGTGAAGAGCTGTGAGTG 1684	
DB	6394	CTGCTGCAAAAACGCTGAGTGG 6373	

RESULT 8  
 US-09-023-715-1/c  
 ; Sequence 1, Application US/09023715  
 ; Patent No. 5998144  
 ;  
 ; GENERAL INFORMATION:  
 ;  
 ; APPLICANT: REFF, Mitchell E.  
 ; APPLICANT: BARNETT, Richard Spence  
 ; APPLICANT: MCLACHLAN, Karen Retta  
 ;  
 ; TITLE OF INVENTION: NOVEL METHOD FOR INTEGRATING GENES AT  
 ; TITLE OF INVENTION: SPECIFIC SITES IN MAMMALIAN CELLS VIA HOMOLOGOUS  
 ; TITLE OF INVENTION: RECOMBINATION AND VECTORS FOR ACCOMPLISHING THE SAME  
 ;  
 ; NUMBER OF SEQUENCES: 2  
 ;  
 ; CORRESPONDENCE ADDRESS:  
 ;  
 ; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.  
 ;  
 ; STREET: P. O. Box 1404  
 ;  
 ; CITY: Alexandria  
 ;  
 ; STATE: Virginia  
 ;  
 ; COUNTRY: United States  
 ;  
 ; ZIP: 22313-1404  
 ;  
 ; COMPUTER READABLE FORM:  
 ;

```

1 MEDIUM TYPE: Floppy disk
2 COMPUTER: IBM PC compatible
3 OPERATING SYSTEM: PC-DOS/MS-DOS
4 SOFTWARE: Patentin Release #1.0, Version #1.30
5 CURRENT APPLICATION DATA:
6 APPLICATION NUMBER: US/09/023, 715
7 FILING DATE:
8 CLASSIFICATION:
9 PRIOR APPLICATION DATA:
10 APPLICATION NUMBER: US 08/819, 866
11 FILING DATE: 14-MAR-1997
12 ATTORNEY/AGENT INFORMATION:
13 NAME: Teskin, Robin L.
14 REGISTRATION NUMBER: 35,030
15 REFERENCE/DOCKET NUMBER: 012712-352
16 TELECOMMUNICATION INFORMATION:
17 TELEPHONE: (703) 836-6620
18 TELEFAX: (703) 836-2021
19 INFORMATION FOR SEQ ID NO: 1:
20 SEQUENCE CHARACTERISTICS:
21 LENGTH: 14683 base pairs
22 TYPE: nucleic acid
23 STRANDEDNESS: single
24 TOPOLOGY: linear
25 MOLECULE TYPE: DNA (genomic)
26
27 US-09-023-715-1

```

Query Match	6.6%;	Score 112;	DB 2;	Length 14683;
Best Local Similarity	48.1%;	Pred. No. 6.7e-24;		
Matches 472; Conservative	0;	Mismatches 450;	Indels 60;	Gaps 3;

QY	757	TACACTCTGGACATCGGAATCAGAACGATCAGCTGGGACCGAGAGAGGCTCTATCTGAAAC	816
Db	7348	TACCCGCTTGCGCGTCGGCATTCCGGTCACTGGCAGTGAAAGGCGAAACAGTTCCTGATTTAAC	72895
QY	817	GGGAAACCCTGTCTTTTGAAGGGCTTTGGAAAGCACCAGGAATTTCCCGTTCTGGGGCAG	876
Db	7288	CACAAACCCTTCTACTTACTGGCTTTGGTCGTATGAAGATGCGGACTTGCGTGGCAAA	72295
QY	877	GGCACCCTTTTATCCATTGATGATAAAGACTTCAACCTTCTGAAGTGATCAACGCGAAT	936
Db	7228	GGATTGATTAACGTGCTGATGCTGCACGACCAACGATTAAATGGACTGGATTAGGGCCAAAC	71655
QY	937	TCTTTCAAGACCTCTCACTATCCTTACAGTGAAGAGTGGCTGATCTTGCCGACAGACTC	996
Db	7168	TCCTACCGTAACCTCGCATTAACCTTACGCTGAAGAGATGCTGACTGGCAGATGAACAT	71099
QY	997	GGAAATCCTTGATAGACGAAGCCCC-----GCACTGTGT	10322
Db	7108	GGCATCGTGTGATTGATGAACCTGCTGTGTGCGCTTTAACTCTCTTAGGCATTGGT	70495
QY	1033	ATCACAAGGTACCACTACAATCCCGAGACTCAGAAGATAGCAGAAGCAAC-----	10833
Db	7048	TTCGAAGCGGGCAACAAGCCGAAGAAGACTGTACAAGAAAGGCGAGTCAACGGGGAAACT	69895
QY	1084	-----ATAAGAGAATGATGCACAGACACAAGAACCATCCAGT	11222
Db	6988	CACCAAGCGCACTTACAGCGGATTAAAGAGCTGATAGCGCGTGAACAAAAACCAACCAAGC	69295
QY	1123	GTGATCATGTGAGTGTGGCGAACGAACAGAGTCCAACCATCCAGACGCGAGGGTTTC	11822
Db	6928	GTGTGATGTGAGTATTGCCAACGAACCGGATACCCGTCGCAAGGTGCACGGGAATAT	68695
QY	1183	TTCAAAGCCCTTATGAGACTGCCAATGAATGATCGAACACGCCCCGTTGTATGTG	12422
Db	6868	TTCCGCGCACTGGCGGAAGACAACGCGTAACTCGACCCGACGCGTCCGATCACTGCGTC	68099
QY	1243	AGCATGATGAGCAGCACCAGACGAGAGAACAGAGACGTGGCGCTGAAGTACTTCGACATC	13022
Db	6808	AATGTAAATGTCTGCGACGCTCACACCCGATACCATCAGCGATCT-----CTTGATGTG	67555
QY	1303	GTCTGTGTGAACAGGTAATAACGCTGTATCATCTATCAAGGAAAGATAGAGAAGACTT	13622



Db 6754 CTGTGCTGAACCGTTATTACGGATGGTATGTCCAAAGCGCGATTGGAAACGGCAGAG 6695  
QY 1363 CAAGCTCTGGAAGAAACATAGAGAGCTCTATGCAAGGACAGAAAGCCCATCTTGTG 1422  
Db 6694 AAGTAATGGAAGAAAGAACTTCTGCGCTGGCAGAGAAATGCAATGACCGATATCATC 6635  
QY 1423 ACAGAAATTCGTCGCGACCGGATGCTGCGATCCATACGATCCACTCAAAATGTTCTCC 1482  
Db 6634 ACCGAATACGCGGTGATACGTTAGCCGGCTGCACTCAATGTAACACCGACATGTGAGT 6575  
QY 1483 GAAGAGTACCAAGAGAGCTCGTTGAAAAGACGATCAGGCTCTTTTGAAGAAAGACTAC 1542  
Db 6574 GAAGAGTATCAGTGTGATGCTGCGATATGATATACCGCGCTTTTGATCGCGTACGCGC 6515  
QY 1543 ATCATCGGAACACAGCTGTGGCTTTGCAAGATTTTAAGACTCTCAGATGTGAGAGA 1602  
Db 6514 GTGCTCGGTGAACAGATATGGAATTTGCGGATTTTGGACCTCGCAAGGCAATTTGCGC 6455  
QY 1603 CCCATTCTCAACCAAGGCTTTTCAAGAGACAGACAAACCCAACTGTTGCTCAT 1662  
Db 6454 GTTGGCGGTAAAGAAAGGATCTTCACTCGCGACCGCAACCGAAGTGGCGGCTTTT 6395  
QY 1663 GTACTGAGAAGACTGTGAGTG 1684  
Db 6394 CTGCTGCAAAAACGCTGACTG 6373

## RESULT 9

US-09-343-485A-1/c

; Sequence 1, Application US/09343485A  
; Patent No. 6413777

; GENERAL INFORMATION:

; APPLICANT: REEF, MITCHELL R.

; APPLICANT: BARNETT, RICHARD S.

; APPLICANT: MCLACHLAN, KAREN R.

; TITLE OF INVENTION: NOVEL METHOD FOR INTEGRATING GENES AT SPECIFIC SITES IN

; TITLE OF INVENTION: MAMMALIAN CELLS VIA HOMOLOGOUS RECOMBINATION AND

; FILE REFERENCE: 037003-0275807

; CURRENT APPLICATION NUMBER: US/09/343,485A

; PRIOR FILING DATE: 1999-06-30

; PRIOR APPLICATION NUMBER: 09/023,715

; PRIOR FILING DATE: 1998-02-13

; PRIOR APPLICATION NUMBER: 08/819,866

; NUMBER OF SEQ ID NOS: 3

; SOFTWARE: Patentln Ver. 2.1

; SEQ ID NO 1

; LENGTH: 14683

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: Synthetic DNA

; OTHER INFORMATION: referred to as "Diamond"

US-09-343-485A-1

Query Match 6.6%; Score 112; DB 4; Length 14683;  
Best Local Similarity 48.1%; Pred. No. 6.7e-24;

Matches 472; Conservative 0; Mismatches 450; Indels 60; Gaps 3;

QY 757 TACACTGTGACATCGAATCAGAACGATCAGCTGGACGAGAGAGGCTCTATCTGAAC 816  
Db 7348 TACCCGCTTCGCGTGGCATCCGCTCAGTGGCAGTGAAGGGCGAAGCTTCGATTAAC 7289  
QY 817 GGAAACCTGTCTTTTGAAGGCTTTGAAAAGACAGAGGAATCCCGTCTTGGGCGAG 876  
Db 7288 CACAAACGCTTCTACTTACTGCTTGGTTCGTCATGAAGATCGGACTTGGCGCAA 7229  
QY 877 GGACCTTTTATCATGATGATAAAGACTTCAACCTTCTGAAGTGAATCAACGGAAT 936  
Db 7228 GGATTGATAACGTGCTGATGTGACAGACCAAGCAATTAATGACTGATTAAGGCGCAAC 7169  
QY 937 TCTTTAGGAGCTCTCACTATCTTACAGTGAAGATGGCTGATCTTGGCGACAGACTC 996

Db 7168 TCCTACCGTACCTCGCATTAACCTTAACGCTGAAGAGATGCTGACTGGGACAGATGAACAT 7109  
QY 997 GGAATCTTGTGATAGCGAAGCCCC-----GCACGTTGGT 1032  
Db 7108 GGATCGTGTGATGATGAAGAACTGCTGTGCTGCGCTTTAACCTCTCTTTAGGCATTTGGT 7049  
QY 1033 ATCAAGAGTACCCTCAATCCCGAGACTCAGAGATAGCAGAAAGCAAC----- 1083  
Db 7048 TTGAAAGCGGCAACAGCCGAAAGAACTGTACACGAAAGAGGACATCAACGGGAAACT 6989  
QY 1084 -----ATAAGAAATGATCGACAGACAGAAAGCAACCATCCAGT 1122  
Db 6988 CACCAAGCGCACTTACAGCGGATTAAGAGCTGATAGCGCGTGACAAAGAACCAACCAAGC 6929  
QY 1123 GTGATCATGTGAGTGTGGCGAAGCAACAGAGTCAACCATCCAGACGGAGGTTTC 1182  
Db 6928 GTGATGATGTGAGTATTTCCCAACGAAACCGATACCCGTCGCAAGGTGACAGGGAATAT 6869  
QY 1183 TTCAAGCCCTTTATGAGACTGCCAATGAATGATCGAAGACAGCCCGTGTGATGTTG 1242  
Db 6868 TTGCGGCCACTGGCGGAAAGCAACGCGTAACCTGACCCGACGCGTCCGATCAGCTGCGTC 6809  
QY 1243 AGCATGATGAGCGCACAGACGAGAGAAACAGAGAGCTGGCGCTGAAGTACTTGCATC 1302  
Db 6808 AATGTAATGTTCTGCGACGCTCACACCGATACCATCAGCATCT-----CTTTGATGTG 6755  
QY 1303 GTCTGTGTAACAGTACTACCGGCTGTGATCATCTATCAGGGAAGATAGAAAGACTT 1362  
Db 6754 CTGTGCTGAACCGCTTATTACGAGATGTTATGTTCCAAAGCGCGATTGGAAACGCGAGAG 6695  
QY 1363 CAAGCTCTGGAAGAAACATAGAGAGCTCTATGCAAGGACAGAAAGCCCATCTTGTG 1422  
Db 6694 AAGTAATGGAAGAAAGAACTTCTGCGCTGGCAGAGAAATGCAATGACCGATATCATC 6635  
QY 1423 ACAGAAATTCGTCGCGACCGGATGCTGCGATCCATACGATCCACTCAAAATGTTCTCC 1482  
Db 6634 ACCGAATACGCGGTGATACGTTAGCCGGCTGCACTCAATGTAACACCGACATGTGAGT 6575  
QY 1483 GAAGAGTACCAAGAGAGCTCGTTGAAAAGACGATCAGGCTCTTTTGAAGAAAGACTAC 1542  
Db 6574 GAAGAGTATCAGTGTGATGCTGCGATATGATATACCGCGCTTTTGATCGCGTACGCGC 6515  
QY 1543 ATCATCGGAACACAGCTGTGGCTTTGCAAGATTTTAAGACTCTCAGATGTGAGAGA 1602  
Db 6514 GTGCTCGGTGAACAGATATGGAATTTGCGGATTTTGGACCTCGCAAGGCAATTTGCGC 6455  
QY 1603 CCCATTCTCAACCAAGGCTTTTCAAGAGACAGACAAACCCAACTGTTGCTCAT 1662  
Db 6454 GTTGGCGGTAAAGAAAGGATCTTCACTCGCGACCGCAACCGAAGTGGCGGCTTTT 6395  
QY 1663 GTACTGAGAAGACTGTGAGTG 1684  
Db 6394 CTGCTGCAAAAACGCTGACTG 6373

## RESULT 10

US-08-723-624-18

; Sequence 18, Application US/08723624

; Patent No. 5861277

; GENERAL INFORMATION:

; APPLICANT: Rose, Alan B.

; APPLICANT: Last, Robert L.

; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR ENHANCING

; TITLE OF INVENTION: THE EXPRESSION OF GENES IN PLANTS

; NUMBER OF SEQUENCES: 21

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: ARNOLD, WHITE &amp; DURKEE

; STREET: P.O. Box 4433

; CITY: Houston

; STATE: TX

; COUNTRY: USA

; ZIP: 77210

```

/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patentin Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/723,624
/ FILING DATE: Concurrently Herewith
/ CLASSIFICATION: 435
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Parker, David L.
/ REGISTRATION NUMBER: 32,165
/ REFERENCE/DOCKET NUMBER: BTIP:002
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 512/418-3000
/ TELEFAX: 512/474-7577
/ INFORMATION FOR SEQ ID NO: 18:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 3035 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/
US-08-723-624-18
```

```

Query Match          6.6%; Score 110.8; DB 2; Length 3035;
Best Local Similarity 47.4%; Pred. No. 5.9e-24;
Matches 465; Conservative 0; Mismatches 457; Indels 60; Gaps 2;
```

```

QY 757 TACACTCTGACATCGGAATCAGATCAGCTGGACGAGAGAAGCTCTATCTGAAC 816
   |||||
DB 2016 TACCCGCTTCGCGTCGGCATCCGGTCAGTGCAGTGAAGGCGAACAAGTTCCTGATTAAC 2075

QY 817 GGGAAACCTGTCTTTTGAAGGGCTTTGGAAGCAAGGAATTCCTGTTGGGGCAG 876
   |||||
DB 2076 CACAACCGTCTTACTTACTGCGTTTGTCTGTCATGAAGATGCGGAAGTTCGCGGCAAA 2135

QY 877 GGCACCTTTTATCCATTGATGATATAAGACTTCAACTTCTGAAGTGATCAACGCGAAT 936
   |||||
DB 2136 GGATTCGATAACGTGCTGATGATGACACGACGACGATTAATGATGATGGGCGCAAC 2195

QY 937 TCTTCAGACCTCTCACTATCTTACAGTGAAGAGTGCTGATCTTGCCGACAGACTC 996
   |||||
DB 2196 TCCTACCGTACCTCGCATTAACCTTACGCTGAAGAGATGCTGACCTGGGACAGATGAACAT 2255

QY 997 GGAATCCTTGATAGACGAAGCCCGCAGCTTGCTGATCACAAGT----- 1042
   |||||
DB 2256 GGCATCGTGTGATGATGAAGATGCTGCTGCTGCTTAACTCTCTTTAGGCAATTGGT 2315

QY 1043 -----ACCACTACAATCCCGAGACT 1062
   |||||
DB 2316 TTCGAAGCGGGCAACAAGCCGAAGAAGTGTACAGCGAAGGCGCATCAACGGGGAAACT 2375

QY 1063 CAGAAGATAGCAGAAGACAACATAGAAGATGATCGACAGACACAAGAACCATCCCACT 1122
   |||||
DB 2376 CAGCAAGCGCACTTACAGCGGATTAAGAAGCTGATAGCGCGTGACAAAAACCAACCAAGC 2435

QY 1123 GTGATCATGTGAGTGTGGCGAAGCAACCAAGATCCCAACCATCCAGCGCGAGGCTTC 1182
   |||||
DB 2436 GTGGTGTGATGTGAGTATTGGCAAGCAACCGGATACCCGTCGCAAGGTGCACGGGAATAT 2495

QY 1183 TTCAAAGCCCTTATAGACTGCGCAATGAATGATCGAACACGCCCCGTTGTATGTTG 1242
   |||||
DB 2496 TTCGCGCCCACTGGCGGAAGCAACCGCTAAACTCGAACCGCGCTCCGATCACTCGCGTC 2555

QY 1243 AGCATGATGAGCGCACCAAGCAGAGAGAACAAGAGAGCTGGCGCTGAAGTACTTGAACATC 1302
   |||||
DB 2556 AATGTAATGTTCTCGCAGCGCTCACACCGATACCATCAGCGATCT-----CTTTGATGTG 2609

QY 1303 GTCTGTGTGAACAGGTAAGTACGGCTGTGATCATCTATCAGGGAAGGATAGAAGAGACTT 1362
   |||||
DB 2610 CTGTGCTGTAACCGTTATTACCGAGTGATGTCCAAAGCGGCGGATTTGAAACGCGCAGAG 2669

QY 1363 CAAAGCTCTGAAAAAGACATAGAAAGAGCTCTATGCAAGGACAGAAAGCCCATCTTTGTC 1422
```

```

DB 2670 AAGTACTGAAAAAGAACTTCGCGCTGCGCAGAGAGAAGAACTGCATCAGCCGATTAATCATC 2729
   |||||
QY 1423 ACAGATTTCGTCGCGGACGCGGATAGCTGCAATCCACTACGATCCAACTCAATGTCTCC 1482
   |||||
DB 2730 ACCGAATACGGCTGTGATACGTTAGCCGGCTGCACTCAATGTATACCCGACATGTGAGT 2789

QY 1483 GAAGAGTACCAAGCAGAGCTCGTTGAAAAAGACGATCAGGCTCCTTTGAAAAAAGACTAC 1542
   |||||
DB 2790 GAAGAGTATCAGTGTGCATGGCTGTGATATGATATCAACCGCTTTGATCCGCTACGCGCC 2849

QY 1543 ATCATCGGAACACAGTGTGGGCTTTGCAAGATTTTAAGACTTCCTCAGATGTGAGAAGA 1602
   |||||
DB 2850 GTGTCGGTGAACAGGATATGGAATTTCCGCAATTTTGCGACCTTCGCAAGGCAATTTGCCG 2909

QY 1603 CCCATTTCAACCAAGGCTTTTCAACAAGACAGACACAACCAACTCGTTGCTCAT 1662
   |||||
DB 2910 GTTGGCGGTAAACAAGGAGATCTTCACTCGCGACCGCAACCGAAGTCGCGGCTTTT 2969

QY 1663 GTAAGTGAAGAGCTGTGAGTG 1684
   |||||
DB 2970 CTGCTGCAAAAACGCTGAGCTG 2991
```

## RESULT 11

```

US-08-630-820-5
; Sequence 5, Application US/08630820
```

```

; Patent No. 6008023
```

```

; GENERAL INFORMATION:
```

```

; APPLICANT: OPPER, Martin
```

```

; APPLICANT: BOSSLET, Klaus
```

```

; APPLICANT: CZECH, Joerg
```

```

; TITLE OF INVENTION: CYTOPLASMIC EXPRESSION OF ANTIBODIES,
```

```

; TITLE OF INVENTION: ANTIBODY FRAGMENTS AND ANTIBODY FRAGMENT FUSION MOLECULES
```

```

; NUMBER OF SEQUENCES: 7
```

```

; CORRESPONDENCE ADDRESS:
```

```

; ADDRESSER: Foley & Lardner
```

```

; STREET: 3000 K Street, N.W., Suite 500
```

```

; CITY: Washington
```

```

; STATE: D.C.
```

```

; COUNTRY: USA
```

```

; ZIP: 20007-5109
```

```

; COMPUTER READABLE FORM:
```

```

; MEDIUM TYPE: Floppy disk
```

```

; COMPUTER: IBM PC compatible
```

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
```

```

; SOFTWARE: Patentin Release #1.0, Version #1.30
```

```

; CURRENT APPLICATION DATA:
```

```

; APPLICATION NUMBER: US/08/630,820
```

```

; FILING DATE: 10-APR-1996
```

```

; PRIOR APPLICATION DATA:
```

```

; APPLICATION NUMBER: DE 19513676.4
```

```

; FILING DATE: 11-APR-1995
```

```

; ATTORNEY/AGENT INFORMATION:
```

```

; NAME: GRANADOS, Patricia D.
```

```

; REGISTRATION NUMBER: 33,683
```

```

; REFERENCE/DOCKET NUMBER: 18748/306
```

```

; TELECOMMUNICATION INFORMATION:
```

```

; TELEPHONE: (202)672-5300
```

```

; TELEFAX: (202)672-5399
```

```

; TELEEX: 904136
```

```

; INFORMATION FOR SEQ ID NO: 5:
```

```

; SEQUENCE CHARACTERISTICS:
```

```

; LENGTH: 3169 base pairs
```

```

; TYPE: nucleic acid
```

```

; STRANDEDNESS: double
```

```

; TOPOLOGY: circular
```

```

; MOLECULE TYPE: cDNA
```

```

; HYPOTHETICAL: NO
```

```

; ANTI-SENSE: NO
```

```

; ORIGINAL SOURCE:
```

```

; ORGANISM: Enterobacteriaceae: Escherichia coli
```

STRAIN: PRAJ210  
IMMEDIATE SOURCE:  
CLONE: pTic99 dicistr. Fab/E.c.-Beta-Gluc  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 3..641  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 666..3162  
US-08-630-820-5

Query Match 6.6%; Score 110.8; DB 3; Length 3169;  
Best Local Similarity 47.4%; Pred. No. 6e-24;  
Matches 465; Conservative 0; Mismatches 457; Indels 60; Gaps 2;

QY 757 TACACTCTGACATCGGAATCAGACGATCAGCTGGGACGAGAGAGGCTCTATCTGAAC 816  
DB 2145 TACCCGCTTCGCGTCGGCATCCGGTCAGTGGCAGTGAAGGGCGAACAAGTCCCTGATTAC 2204  
QY 817 GGGAACTGTCTTTTGAAGGGCTTGAAGAAGCAGGAATTCCCGCTTCTGGGGCAG 876  
DB 2205 CACAAACCGTCTTACTTACTGCTTGTGCTGATGAAGATGGGACTTACTGGCAA 2264  
QY 877 GGCACCTTTTATCATTTGATGATAAAGACTTCAACCTTCTGAAGTGATCAACCGAAT 936  
DB 2265 GGATTCGATAACGTGCTGATGGTGACAGACACGACGATTAATGAGCTGATTGGGCCAAC 2324  
QY 937 TCTTCAGAGACCTCTCACTATCCTTACAGTGAAGAGTGGCTGGATCTTGCCGACAGACTC 996  
DB 2325 TCCTACCGTACCTCGCATTTACCTTACGCTGAAGAGATGCTGCACTGGGCGAGATGAACAT 2384  
QY 997 GGAATCCTTGTGATAGAGAGCCCGCAGCTGTGATCAAGGT----- 1042  
DB 2385 GGCATCGTGTGATGATGAAGCTGCTGCTGCTTAACTCTCTTAGGCATTGCT 2444  
QY 1043 -----ACCACTACAATCCCGAGACT 1062  
DB 2445 TTCGAAGCGGCAACAAGCCGAAGAAGTGTACAGGAGGCGACTCAACGGGGAAC 2504  
QY 1063 CAGAAATAGCAGAGACAACAATAAGAAATGATCGACAGACACAAGAACCATCCAGT 1122  
DB 2505 CAGCAAGCGCACTTACAGCGATTAAGAGCTGATAGCGCGTGAACAACCAACCAAGC 2564  
QY 1123 GTGATCATGTGAGTGTGGCGAAGCAGAGTCCAACCATCCAGACGCGAGGTTTC 1182  
DB 2565 GTGTGATGTGAGTATTGCGAAGCAGATACCGGTCCGCAAGTGCACGCGGAATAT 2624  
QY 1183 TTCAAAGCCCTTATGAGACTGCCAATGAATGATCGAACAGCCCGCTTGTGATG 1242  
DB 2625 TTCGCGCCTGCGGAGCAAGCGTAACCTCGACCGCGCTCCGATCACTGCGTC 2684  
QY 1243 AGCATGATGAGCAGCAGCAGAGAGAGAGAGAGAGAGTGGCGCTGAAGTACTTGACATC 1302  
DB 2685 AATGTAAATGTTCTGCGACGCTCACACCGATACCATCAGCATCT-----CTTGATG 2738  
QY 1303 GTCTGTGACAGAGTACTACGCTGTGATCTATCAGGGAAGATAGAAGAGACTT 1362  
DB 2739 CTGTGCTGAACCGTTATTACGATGTGTCCAAAGCGCGATTTGAAAACGCGAG 2798  
QY 1363 CAAGCTCTGAAAAAGACATAGAAGAGCTGTATGCAAGGCAAGAAAGCCCATCTTGT 1422  
DB 2799 AAGTACTGAAAAAGAACTTCTGCGCTGCGAGAGAACTGCATCAGCCGATTATCTC 2858  
QY 1423 ACAGAAATCGGTGCGAGCGATAGCTGCGATCCACTAGATCCACTCAATGTTCTCC 1482  
DB 2859 ACCGAATACGGCGTGAATAGTACCGGCGCTGCACTCAATGTACACGACGATGTGAGT 2918  
QY 1483 GAAGAGTACCAAGCAGAGCTGTTGAAAAGACGATCAGGCTCTTTGAAAAGACTAC 1542  
DB 2919 GAAGAGTATCAGTGTGATGCTGATGTATATCACCAGGCTTTGATCGCGTACAGGCC 2978  
QY 1543 ATCATCGAACAACAGTGTGGGCTTTGAGATTTTAAGACTCTCAGAAATGTGAAGA 1602

DB 2979 GTGTCGGTGAACAGGTATGGAATTTCCCGGATTTTGGCACTCGCAAGCATATTGCCG 3038  
QY 1603 CCCATTCTCAACACAGAGGCTGTTTCAAGAGACAGACAAACCAACTGCTGCTCAT 1662  
DB 3039 GTTGGCGGTAAACAGAAAGGATCTTCACTCGGACCGCAACCGAAGTCCGGGCTTTT 3098  
QY 1663 GTACTGAGAAGACTGTGAGCTG 1684  
DB 3099 CTGCTGCAAAAACGCTGAGCTG 3120

## RESULT 12

US-09-273-453-5

Sequence 5, Application US/09273453

Patent No. 6602688

GENERAL INFORMATION:

APPLICANT: OPPER, Martin

BOSSLET, Klaus

CZECH, Joerg

TITLE OF INVENTION: CYTOPLASMIC EXPRESSION OF ANTIBODIES,

ANTIBODY FRAGMENTS AND ANTIBODY FRAGMENT FUSION MOLECULES

IN E. COLI

NUMBER OF SEQUENCES: 7

CORRESPONDENCE ADDRESS:

ADDRESSEE: Foley &amp; Lardner

STREET: 3000 K Street, N.W., Suite 500

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20007-5109

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/273,453

FILING DATE: 22-Mar-1999

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/630,820

FILING DATE: &lt;Unknown&gt;

ATTORNEY/AGENT INFORMATION:

NAME: GRANADOS, Patricia D.

REGISTRATION NUMBER: 33,683

REFERENCE/DOCKET NUMBER: 18748/306

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 672-5300

TELEFAX: (202) 672-5399

TELEX: 904136

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 3169 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: circular

MOLECULE TYPE: cdna

HYPOTHETICAL: NO

ANTI-SENSE: NO

ORIGINAL SOURCE:

ORGANISM: Enterobacteriaceae: Escherichia coli

STRAIN: PRAJ210

IMMEDIATE SOURCE:

CLONE: pTic99 dicistr. Fab/E.c.-Beta-Gluc

FEATURE:

NAME/KEY: CDS

LOCATION: 3..641

FEATURE:

NAME/KEY: CDS

LOCATION: 666..3162

SEQUENCE DESCRIPTION: SEQ ID NO: 5:

US-09-273-453-5

Query Match 6.6%; Score 110.8; DB 4; Length 3169;

Best Local Similarity 47.4%; Pred. No. 6e-24;  
Matches 465; Conservative 0; Mismatches 457; Indels 60; Gaps 2;

QY	757	TACACTCTGGACATCGGAATCAGAACGATCAGCTGGGACGAGAAAGGCTCTATCTGAAC	816
Db	2145	TACCCTTCGCGTCGGCATCCGCTCAGTGGCAGTGAAGGGCGAAACAGTTCCTGATTAAAC	2204
QY	817	GGGAAACCTGTCTTTTGAAGGCGTTTGGAAAGCAGAGAAATTCCTCGTTGCGGGCAG	876
Db	2205	CACAAACCGTTCTTACTTTACTGCGTTTGGTCGTTCATGAAGATCGGACTTACGTGGCAAA	2264
QY	877	GGCACCCTTTTATCCATTGATGATTAAGACTTCAACCTTCTGAAGTGATCAACGGAAT	936
Db	2265	GGATTTCGATTAACGTGCTGATGGTGCACGACCAAGCATTAAATGACCTGATTGGGCCAAC	2324
QY	937	TCTTTCAGGACCTCTCACTATCCTTACAGTGAAGAGTGGCTGGATCTTGCGCAGACTC	996
Db	2325	TCTTACCGTACCTCGCATTACCTCTTACGCTGAAGAGATGCTGCACCTGGCAGATGAACAT	2384
QY	997	GGATCCTTGTGATAGACGAAGCCCGCACGTTGGTATCACAAGT-----	1042
Db	2385	GGCATCGTGTGATGATGAACAGCTGCTGTCGGCTTAACTCTCTTAGGCATTGGT	2444
QY	1043	-----ACCACTACAATCCGAGACT	1062
Db	2445	TTCCGAAGCGGGCAACAAGCCGAAGAACTGTACAGCGAAGAGGCGAGTCACCGGGGAACT	2504
QY	1063	CAGAAAGTAGCAGAGAACAACATAAGAAATGATCGACAGACACAAGAACCATCCAGT	1122
Db	2505	CAGCAAGCGCACTTACAGCGGATTAAGAGCTGATAGCGCGTGCACAAAAACCAACCAAGC	2564
QY	1123	GTCATCATGTGAGTGTGGCAACGAACCAAGATCCAACATCCAGACGCGGAGGTTTC	1182
Db	2565	GTCGTGATGTGAGATATTGCCAAGAACCGGATACCCGTCGCGAAGGTGCACGGGAATAT	2624
QY	1183	TTCAAAGCCCTTTATGAGACTGCCAATGAATGATCGAACACGCCCGTTGTCATGTTG	1242
Db	2625	TTCCGCCCACTGGCGGAAGCAACCGCTAACTCGAACCCGACGCGTCCGATCCTGCGTC	2684
QY	1243	AGCATGATGGACCGCACCAAGCAGAGAACACAGACGTCGCGCTGAAGTACTTCGACATC	1302
Db	2685	AATGTAATGTTCTGCGACGCTCACACCGATACCATCAGCGATCT-----CTTTGATGTG	2738
QY	1303	GTCGTGTGAACAGGTACTACGCGCTGTGATCATCTATCAGGGAAAGATGAGAAGGACTT	1362
Db	2739	CTGTGCTGAAACCGTTATTACCGATGGTATGTCTCAAAAGCGGCGATTGGAAACGGCAGAG	2798
QY	1363	CAAGCTCTGGAAAAAGACATAGAAAGCTCTATGCAAGGCACAGAAAGCCCATCTTTGTC	1422
Db	2799	AAGGTACTGGAAAAAGAACTTCTGGCCTGGCAGAGAAACTGCATCAGCCGATTATCATC	2858
QY	1423	ACAGAATTGCGTGGCGACGCGATAGCTGGCATCCACTACGATCCACTCAAAATGTTCTCC	1482
Db	2859	ACCGAATACGGCGTGATACGTTAGCCGGGCTGCACCTCAATGTACACCGACATGTGAGT	2918
QY	1483	GAAAGTACCAAGCAGAGCTCGTTGAAAAGACGATCAGGCTCCTTTTGA AAAAGACTAC	1542
Db	2919	GAAAGTATCAGTGTGCATGCGCTGATATGTATCAACCGCTTTTGATCGCGTCAGCGCC	2978
QY	1543	ATCATCGGAACACAGTGTGGGCTTTGCAGATTTTAAAGCTCCTCAGAATGTGAGAAGA	1602
Db	2979	GTCGTGCGTGAACAGTATGAAATTCGCCGATTTTGCAGCTTCGCAAGGCATATTGCGC	3038
QY	1603	CCGATTCTCAACCAACAAGGTTGTTTACAAAGAGACAGACAAACCAACTCGTTGCTCAT	1662
Db	3039	GTTGGCGGTAAACAAGAAAGGATCTTCACTCGGCAACCGAAACCGAAGTCGGCGGCTTTT	3098
QY	1663	GTA CTGAGAAGACTGTGAGTG 1684	
Db	3099	CTGCTGC AAAAACGCTGACTG 3120	

US-08-723-624-19  
; Sequence 19, Application US/08723624

; GENERAL INFORMATION:

APPLICANT: Last, Robert L.

**TITLE OF INVENTION: THE EXPRESSION OF GENES IN PLANTS**

CORRESPONDENCE ADDRESS:

STREET: P.O. Box 4433

STATE: TX

COMPUTER READABLE FORM:

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS

```

; SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

```

APPLIATION NUMBER:  
FILING DATE: CONCUR

CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:

NAME: FALVEL, DAVID L.  
REGISTRATION NUMBER: 32,165

TELECOMMUNICATION INFORMATION:

TELEFAX: 512/474-7577

SEQUENCE CHARACTERISTICS:

TYPE: nucleic acid

US-08-723-624-19

Query Match	6.6%;	Score 110.8;	DB 2;	Length 3824;
Best Local Similarity	47.4%;	Pred. No. 6.8e-24;		
Matches 465; Conservative	0;	Mismatches 457;	Indels 60;	Gaps 2;

OY	757	TACACTTGGACATCGGAATCAGAA	CGATCAGCTGGGACGAGAAGGCTCTATCTGAAC	816	
Db	2805	TACCCGCTTCGCGTCGGCATTCGGT	CAGTGGCAGTGAAGGGCGAACAGTTCTCTGATTAAC	2864	
OY	817	GGGAAACCTGTCTTTTGAAGGGCT	TTGGAAAGCAGCAGGAATTC	876	
Db	2865	CACAAACCGTTCTACTTTACTTT	AGCTTTGGTCGTGTCATGAAGATGCGGACTTGGCTGGCAAA	2924	
OY	877	GGCACCTTTTATCCATTGATGAT	TAAGAAGCTCAACCTTCTGAAGTGATCAACGCGAAT	936	
Db	2925	GGATTGATTAACGTGCTGATG	GTGCACGACCAAGCATTAATGAGCTGGAATGGGGCCAAAC	2984	
OY	937	TCTTTACGACCTCTCACTATCCT	TACAGTGAAGAGTGGCTGATCTTGCCGACAGACTC	996	
Db	2985	TCCTACCGTAACCTCGCATTAAC	CTTAACGTGAAGAGATGCTCGACTGGGACATGAACAT	3044	
OY	997	GGAAATCCTTGATAGACGAAGCC	CGCAGCTTGATATCAAAAGT-----	1042	
Db	3045	GGCATCGTGTGATTGATGA	AACTGCTGTGCGCTTTAACTCTCTTTAAGGCATTGGT	3104	
OY	1043	-----	-----ACCAC	TAATCCCGAGACT	1062
Db	3105	TTCGAAGCGGGCAACAAAGCC	GAAGAAGTGTACAGCGAAGAGGCAGTCAACGGGGAAACT	3164	
OY	1063	CAGAAAGATAGCAGAAGACA	CATTAAAGAAGATGATGCAGACACAAAGAACCATCCCAGT	1122	
Db	3165	CAGCAAGCGCACTTACAGCG	CAATTAAAGAGCTGATAGCGCGTGACAAAAAACCAACCCAAAGC	3224	
OY	1123	GTGATCATGTGAGGTGTGGCG	AAACGAAACGAACTCCAAACATCCAGACCGGAGGGTTTC	1182	

```
Db 3225 GTGTGATGTGAGTATTTGCCAACGACCGGATACCCGTCGCAAGGTGCACCGGAATAT 3284
QY 1183 TTCAAAGCCCTTTATGAGACTGCGCATGAAATGATCGAACACGCCCCGTTGTGATGTG 1242
Db 3285 TTGGGCGCACTGGCGGAAAGCAACGCGTAAACTGACCCGACGCGTCCGATCACTGCGTC 3344
QY 1243 AGCATGATGACGACCAAGACGAGAGAAACAAGAGACGTGCGCTGAAGTACTTGCACATC 1302
Db 3345 AATGTAATGTTCTGGGACGCTCACACCGATACATCAGCGATCT-----CTTTGATGTG 3398
QY 1303 GTCTGTGTGAACAGGTAATACGCGCTGTACATCTATCAGGAAAGATAGAAGAGACTT 1362
Db 3399 CTGTGCTGAACCGTTATTAACGAGTGTATGTCCAAACGCGCATTTGAAACGGCAGAG 3458
QY 1363 CAAGCTCTGAAAAAGACATAGAAGAGCTTATGCAAGGACAGAAAGCCCATCTTTGTC 1422
Db 3459 AAGGTAATGAAAAAGAACTTCTGCGCTGCGAGAGAACTGCATCAGCCGATATCATC 3518
QY 1423 ACAGAAATTCGTCGGGACGCGATAGCTGCGATCCATCAATCCACTCAATGTTCTCC 1482
Db 3519 ACCGAATACGCGCTGATACGTTAGCCGCGCTCACTCAATGTACACCGACATGTGAGT 3578
QY 1483 GAAGAGTACCAAGCAGAGCTCGTTGAAAAAGACGATCAGGCTCTTTTGAAGAAAGACTAC 1542
Db 3579 GAAGAGTATCAGTGTGATGCTGATGATATGATACCGCGCTTTGATCGCGTCAGCGCC 3638
QY 1543 ATCATCGGAACACGCTGTGGGCTTTGAGATTTTAAGACTCTCTCAAGATGTGAGAGA 1602
Db 3639 GTGCTCGGTGAACAGGTATGGAATTTCCCGGATTTTGCAGCTCGCAAGGATATTGCGC 3698
QY 1603 CCCATTCTCAACCAAGGCTGTTTTCACAAGAGACAGACCAACCAACTGCTGCTCAT 1662
Db 3699 GTTGGCGGTAAAGAAAGGATCTTCACTGCGACCGCAACCGAAGTCGCGGCTTTT 3758
QY 1663 GTACTGAGAAGACTGTGAGTG 1684
Db 3759 CTGCTGCAAAAACGCTGACTG 3780
```

## RESULT 14

```
US-09-893-525-36
; Sequence 36, Application US/09893525
; Patent No. 6753167
; GENERAL INFORMATION:
; APPLICANT: Moloney, Maurice M.
; APPLICANT: Van Rooijen, Gijb
; TITLE OF INVENTION: Preparation of Heterologous Proteins on Oil Bodies
; FILE REFERENCE: 9369-172
; CURRENT APPLICATION NUMBER: US/09/893,525
; CURRENT FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 09/210,843
; PRIOR FILING DATE: 1998-12-15
; PRIOR APPLICATION NUMBER: US 08/846,021
; PRIOR FILING DATE: 1997-04-25
; PRIOR APPLICATION NUMBER: US 08/366,783
; PRIOR FILING DATE: 1994-12-30
; PRIOR APPLICATION NUMBER: US 08/142,418
; PRIOR FILING DATE: 1993-11-16
; PRIOR APPLICATION NUMBER: US 07/659,835
; PRIOR FILING DATE: 1991-02-22
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 36
; LENGTH: 4652
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Phas-GUS-phas
; NAME/KEY: CDS
; LOCATION: (1548)..(3359)
; OTHER INFORMATION:
US-09-893-525-36
```

```
Query Match 6.6%; Score 110.8; DB 4; Length 4652;
Best Local Similarity 47.4%; Pred. No. 7.7e-24;
Matches 465; Conservative 0; Mismatches 457; Indels 60; Gaps 2;

QY 757 TACACTCTGACATCGGAATCAGAACGATCAGCTGGACGAGAGAGGCTCTATCTGAC 816
Db 2343 TACCCGCTTCGCGTCGATCCGATGAGTGGACGTGAAGGCCCAAGTTCCTGATTAAC 2402
QY 817 GGGAAACCTGTCTTTTGAAGGCTTTGAAAGCAGAGGAATTCCTGCTGGGCGAG 876
Db 2403 CACAAACCGTTCTACTTTAAGTGTGCTGCTGCTGATGAAGATGCGCACTTACGTGCAAA 2462
QY 877 GGCACCTTTTATCCATTTGATGATTAAGACTTCAACCTTCTGAAGTGGATCAACCGCAAT 936
Db 2463 GGATTGATTAACGTGCTGATGTTGTCACGACCAACGCACTTAATGACTGGAATTGGGCGCAAC 2522
QY 937 TCTTTACGAGACCTCTCACTATCTCTTAAGTGAAGAGTGGCTGATCTTCCGACAGACTC 996
Db 2523 TCCTACCGTACCTCGCATTAACCTTTACGCTGAAGAGATGCTCGACTGGGCAAGATGACAT 2582
QY 997 GGAATCTTGTGATAGACGAAGCCCGCACGTTGGTATCAACAAGT----- 1042
Db 2583 GGCATGTGTGATGATGAAGAACTGCTGCTGCTGCTTTTCCGCTCTTTAGGCAATTGGT 2642
QY 1043 -----ACCACTCAATCCGAGACT 1062
Db 2643 TTGGAAGCGGGCAACAGCCGAAGAACTGTACAGCGAAGAGGCACTCAACGGGGAACCT 2702
QY 1063 CAGAAATAGCAGAAACACATAGAAGATGATCGACAGACCAAGAACCATCCAGT 1122
Db 2703 CAGCAAGCGCACTTACAGCGGATTAAGAGCTGATAGCGCGTGACAAAAACCAACCAAGC 2762
QY 1123 GTGATCATGTGAGTGTGGCGAAGCAACGAGTCCCAACCATCCAGACCGGAGGTTTC 1182
Db 2763 GTGTGATGTGAGTATGTCACAGAACCGGATACCCGTCGCAAGGTGACCGGGAATAT 2822
QY 1183 TTCAAAGCCCTTTATGAGACTGCGCAATGAATGATCGAACACGCCCTTGTGATGTG 1242
Db 2823 TTGCGGCCACTGGCGGAAGCAACCGCTAAACTGCAACCGCGCTCCGATCACTGCGTC 2882
QY 1243 AGCATGATGACGACCAAGACGAGAGACAGACAGAGAGCTGCGCTGAAGTACTTGCACATC 1302
Db 2883 AATGTAATGTTCTGCGACGCTCACACGATACCATCAGCGATCT-----CTTTGATGTG 2936
QY 1303 GTCTGTGTAACAGGTACTACGCGCTGTACATCTATCAGGAAAGATAGAAGAGACTT 1362
Db 2937 CTGTGCTGAACCGTTATTAACGATGTATGTCCAAAGCGGCGATTTGGAACGCGCAGAG 2996
QY 1363 CAAGCTCTGAAAAAGACATAGAAGAGCTTATGCAAGGACAGAAAGCCCATCTTTGTC 1422
Db 2997 AAGTACTGAAAAAAGAACTTCTGCGCTGGCAGAGAAACTGCATCAAGCCGATTAATCATC 3056
QY 1423 ACAGAAATTCGTCGGGACGCGGATAGCTGGCATCCACTACGATCCACTCAATGTTCTCC 1482
Db 3057 ACCGAATACGCGGTGATACGTTAGCCGCGCTGCATCAATGTACACCGCATGTGAGT 3116
QY 1483 GAAGAGTACCAAGCAGAGCTCGTTGAAGACGATCAGGCTCTTTTGAAGAAAGACTAC 1542
Db 3117 GAAGAGTATCAGTGTGATGCTGATATGATCAACCGCGCTTTGATCGGCTCAGCGCC 3176
QY 1543 ATCATCGGAACACGTCGTGGGCTTTGAGATTTTAAGACTCTCAGAGATGTGAGAGA 1602
Db 3177 GTGCTCGGTGAACAGGTATGAAATTTCCGCGATTTTGGCAGCTCGCAAGGATATTGCGC 3236
QY 1603 CCCATTCTCAACCAAGGCTTTTTCACAAGAGACAGACCAACCAACTCTGCTCAT 1662
Db 3237 GTTGGCGGTAAAGAAAGGATCTTCACTCGGACCGCAAAACGAAAGTCGGCGGCTTTT 3296
QY 1663 GTACTGAGAAGACTGTGAGTG 1684
Db 3297 CTGCTGCAAAAACGCTGACTG 3318
```



RESULT 15  
US-09-893-525-41  
Sequence 41, Application US/09893525  
Patent No. 6753167  
GENERAL INFORMATION:  
APPLICANT: Moloney, Maurice M.  
APPLICANT: Van Rooijen, Gijb  
TITLE OF INVENTION: Preparation of Heterologous Proteins on Oil Bodies  
FILE REFERENCE: 9369-172  
CURRENT APPLICATION NUMBER: US/09/893,525  
CURRENT FILING DATE: 2001-06-29  
PRIOR APPLICATION NUMBER: US 09/210,843  
PRIOR FILING DATE: 1998-12-15  
PRIOR APPLICATION NUMBER: US 08/846,021  
PRIOR FILING DATE: 1997-04-25  
PRIOR APPLICATION NUMBER: US 08/366,783  
PRIOR FILING DATE: 1994-12-30  
PRIOR APPLICATION NUMBER: US 08/142,418  
PRIOR FILING DATE: 1993-11-16  
PRIOR APPLICATION NUMBER: US 07/659,835  
PRIOR FILING DATE: 1991-02-22  
NUMBER OF SEQ ID NOS: 42  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 41  
LENGTH: 5390  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: phas-caleo-GUS-phas  
NAME/KEY: CDS  
LOCATION: (1548)..(4097)  
OTHER INFORMATION:  
US-09-893-525-41

Query Match 6.6%; Score 110.8; DB 4; Length 5390;  
Best Local Similarity 47.4%; Pred. No. 8.5e-24;  
Matches 465; Conservative 0; Mismatches 457; Indels 60; Gaps 2;

QY 757 TACACTCTGGACATCGGAATCAGAACGATCAGCTGGGACGAGAAAGGCTCTATCTGAAC 816  
DB 3081 TACCCGCTTCGGCTCGGCATCCGGTCAGTGAGGAGGCGCAACAGTTCTTGATTAAC 3140  
QY 817 GGGAAACCTGCTTTTGAAGGGCTTTGAAAGCAGAGAAATCCCGTTCTGGGGCAG 876  
DB 3141 CACAAACCGTTTACTTACTGCTTTGGTCTCATGAAGATGCGGACTTACGTGGCAA 3200  
QY 877 GGCACCTTTTATCATGTATGATAAAGACTTCACTTCTGAAGTGAATCAACCGAAT 936  
DB 3201 GGAATCGATAACGTGCTGATGTGACAGACGACGATTAATGACTGATTGGGCCAAC 3260  
QY 937 TCTTCAGGACCTCTCACTATCTTACAGTGAAGAGTGGCTGATCTTGCCGACAGACTC 996  
DB 3261 TCCTACCGTACCTCGCATTAACCTTAACGCTGAAGAGATGCTGACTGGGAGATGAACAT 3320  
QY 997 GGAATCCTTGTGATAGACGAAGCCCGCACGTTGGTATCAAAAGT----- 1042  
DB 3321 GGCATCGTGTGATGATGAACCTGCTGCTGCTTTTCTCTTTAGCATTTGGT 3380  
QY 1043 -----ACCACTACAATCCGAGACT 1062  
DB 3381 TTCGAAGCGGCAACAAGCCGAAGAAGTGTACAGCAAGAGGCAAGTCAACGGGAAACT 3440  
QY 1063 CAGAAGATAGCAGAAAGACAATAGAAATGATGACAGACAGACAAGAACCATCCAGT 1122  
DB 3441 CAGCAAGCGCACTTAAGGCGATTAAGAGCTGATAGCGCGTGAACAAAAACCAAGC 3500  
QY 1123 GTGATCATGTGAGTGTGGGAAAGCAACGAGTCCAAACCATCCAGCGGAGGTTTC 1182  
DB 3501 GTGGTGTATGTGATATTGCCAAGAACCGATACCCGTCGCAAGGTGACGGGAATAT 3560  
QY 1183 TTCAAAGCCCTTATGAGACTGCAATGAATGATGCAACAGCCCGTTGTATGTG 1242

DB 3561 TTCGCGCCACTGGCGGAAGCAACCGGTAAACTCGAACCCGACCGCTCCGATCACTGCGTC 3620  
QY 1243 AGCATGATGACGCAACAGACGAGAAACAAGACGCTGGCGCTGAAGTACTTCGATC 1302  
DB 3621 AATGTATGTTCTGCGACGCTCAACCGATACCATCAAGCATCT-----CTTTGATGTG 3674  
QY 1303 GTCTGTGTAACAGTACTACGCTGTGATCATCTATCAGGAAAGATGAAGAAGACTT 1362  
DB 3675 CTGTGCTGTAACCGTTTATTACGATGTGATGTCAAAAGCGGCAATTGGAAACGCAAG 3734  
QY 1363 CAAGCTGTGAAAAAGACATAGAGAGCTCTATGCAAGGCAAGAAAGCCCATCTTGTG 1422  
DB 3735 AAGGTACTGAAAAAGAACTTCTGGCTGCGAGAGAAACTGCATCAAGCCGATTAATC 3794  
QY 1423 ACAGAAATTCGGTGGGACGCGATAGCTGCAATCCATCAACCTCAATGTTCTCC 1482  
DB 3795 ACCGAATACGCGGTGATACGTTAGCCGGCTGCACTCAATGTACACCGCATGTGAGT 3854  
QY 1483 GAAGATACCAAGCAGAGCTCGTTGAAAAAGACGATCAGGCTCTTTGAAAAAGACTAC 1542  
DB 3855 GAAGATATCAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3914  
QY 1543 ATCATCGAACAACAGTGTGGGCTTTGACAGATTTTAAGACTCTCAGAAATGTGAAGA 1602  
DB 3915 GTGCTCGTGAACAGTATGAAATTTGCCGATTTTGCGACCTGCAAGCATATGCGC 3974  
QY 1603 CCCATTTCTCAACCAAGAGGTGTTTCAACAAGACAGACAAACCAAACTCGTTGCTCAT 1662  
DB 3975 GTTGGCGGTAAACAAGAAAGGATCTTCACTCGCAACGCAACCAAGTGTGCGGCTTTT 4034  
QY 1663 GTACTGAGAAGACTGTGAGTG 1684  
DB 4035 CTGCTGCAAAAACGCTGACTG 4056

Search completed: January 24, 2005, 07:33:15  
Job time : 170 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 24, 2005, 03:41:50 ; Search time 958 Seconds  
(without alignments)  
10130.277 Million cell updates/sec

Title: US-09-936-759-14  
Perfect score: 1689  
Sequence: 1 atggttaagaccgcaacgaaa.....gaagactgtgagtgagctt 1689

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4300275 seqs, 2872944193 residues

Total number of hits satisfying chosen parameters: 8600550

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications NA:\*

1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq:\*  
10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq:\*  
12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*  
13: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq:\*  
14: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq:\*  
15: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq:\*  
16: /cgn2\_6/ptodata/2/pubpna/US10D\_PUBCOMB.seq:\*  
17: /cgn2\_6/ptodata/2/pubpna/US10E\_PUBCOMB.seq:\*  
18: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*  
19: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq:\*  
20: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*  
21: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	1688	99.9	1689	15	US-10-364-649-14 Sequence 14, Appl
2	1688	99.9	1692	15	US-10-369-493-26682 Sequence 26682, A
3	174	10.3	2169	14	US-10-136-841-5 Sequence 5, Appl
4	174	10.3	2169	16	US-10-272-531A-5 Sequence 5, Appl
5	174	10.3	2169	16	US-10-272-483A-5 Sequence 5, Appl
6	172.4	10.2	1956	15	US-10-421-175-1 Sequence 1, Appl
7	172.4	10.2	2191	15	US-10-388-360-343 Sequence 343, App
8	172.4	10.2	2191	15	US-10-429-802-16 Sequence 16, Appl
9	172.4	10.2	2191	16	US-10-430-503-7 Sequence 7, Appl
10	172.4	10.2	2191	18	US-10-335-053-51 Sequence 51, Appl
11	166.8	9.9	2472	16	US-10-388-934-225 Sequence 225, App
12	159.6	9.4	1947	15	US-10-421-175-3 Sequence 3, Appl

13	141.4	8.4	1887	15	US-10-120-145-7	Sequence 7, Appl
14	141.4	8.4	3651	15	US-10-322-656-45	Sequence 45, Appl
15	141.4	8.4	4084	15	US-10-322-656-48	Sequence 48, Appl
16	130.4	7.7	1888	15	US-10-364-649-27	Sequence 27, Appl
17	114	6.7	3451	15	US-10-161-403-108	Sequence 108, App
18	114	6.7	3451	18	US-10-161-408-20	Sequence 20, Appl
19	114	6.7	14627	15	US-10-161-403-109	Sequence 109, App
20	114	6.7	14627	18	US-10-161-408-21	Sequence 21, Appl
21	112	6.6	14683	13	US-10-109-853-1	Sequence 1, Appl
22	112	6.6	14683	17	US-10-817-950-1	Sequence 1, Appl
23	110.8	6.6	1809	17	US-10-432-777-16	Sequence 16, Appl
24	110.8	6.6	1812	15	US-10-161-403-105	Sequence 105, App
25	110.8	6.6	1812	15	US-10-369-493-24529	Sequence 24529, A
26	110.8	6.6	1812	18	US-10-161-408-16	Sequence 16, Appl
27	110.8	6.6	1812	18	US-10-149-533A-33	Sequence 33, Appl
28	110.8	6.6	2000	9	US-09-887-576-582	Sequence 582, App
29	110.8	6.6	2001	16	US-10-239-907A-38	Sequence 38, Appl
30	110.8	6.6	2001	17	US-10-332-406A-21	Sequence 21, Appl
31	110.8	6.6	2001	17	US-10-800-161-29	Sequence 29, Appl
32	110.8	6.6	2186	18	US-10-755-328-2	Sequence 2, Appl
33	110.8	6.6	2725	16	US-10-239-907A-43	Sequence 43, Appl
34	110.8	6.6	2730	16	US-10-239-907A-41	Sequence 41, Appl
35	110.8	6.6	4072	16	US-10-239-907A-49	Sequence 49, Appl
36	110.8	6.6	4341	16	US-10-239-907A-47	Sequence 47, Appl
37	110.8	6.6	4652	10	US-09-893-525-36	Sequence 36, Appl
38	110.8	6.6	4652	18	US-10-763-380-36	Sequence 36, Appl
39	110.8	6.6	5390	10	US-09-893-525-41	Sequence 41, Appl
40	110.8	6.6	5390	18	US-10-763-380-41	Sequence 41, Appl
41	110.8	6.6	5418	10	US-09-893-525-38	Sequence 38, Appl
42	110.8	6.6	5418	18	US-10-763-380-38	Sequence 38, Appl
43	110.8	6.6	5898	18	US-10-640-422-160	Sequence 160, App
44	110.8	6.6	8654	10	US-09-845-064-11	Sequence 11, Appl
45	110.8	6.6	8654	10	US-09-845-064-14	Sequence 14, Appl

ALIGNMENTS

RESULT 1  
US-10-364-649-14  
; Sequence 14, Application US/10364649  
; Publication No. US20030229921A1  
; GENERAL INFORMATION:  
; APPLICANT: Richard A. Jefferson and Jorge E. Mayer  
; TITLE OF INVENTION: MICROBIAL B-GLUCURONIDASE GENES, GENE  
; FILE REFERENCE: 190106.405C1  
; CURRENT APPLICATION NUMBER: US/10/364,649  
; PRIOR FILING DATE: 2003-02-12  
; PRIOR APPLICATION NUMBER: 10/364,649  
; PRIOR FILING DATE: 2003-02-12  
; PRIOR APPLICATION NUMBER: US 09/270,957  
; NUMBER OF SEQ ID NOS: 112  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 14  
; LENGTH: 1689  
; TYPE: DNA  
; ORGANISM: Thermotoga maritima  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(1689)  
; OTHER INFORMATION: n = A,T,C or G  
US-10-364-649-14

Query Match 99.9%; Score 1688; DB 15; Length 1689;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1689; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCTAAGACCGCAAGAAAGAGATTATCTTATCTTGAATGAGTTGGAAT 60  
Db 1 ATGCTAAGACCGCAAGAAAGAGATTATCTTATCTTGAATGAGTTGGAAT 60

QY 61 CTTGAAGTAACGAGAAAGACAGACCAATCGCCGTTCTGGAAGCTGGAATGACAGTAC 120  
DB 61 CTTGAAGTAACGAGAAAGACAGACCAATCGCCGTTCTGGAAGCTGGAATGACAGTAC 120  
QY 121 CAGGATCTGTGCTACGAAAGAGGACCTTCACTTCAAAAACACCTTCTACGTTCCGAAG 180  
DB 121 CAGGATCTGTGCTACGAAAGAGGACCTTCACTTCAAAAACACCTTCTACGTTCCGAAG 180  
QY 181 NAACTTTCAAAAAACACATCAGACTTTACTTTGCTGCGGTGAACAACGACTGCGAGGTC 240  
DB 181 NAACTTTCAAAAAACACATCAGACTTTACTTTGCTGCGGTGAACAACGACTGCGAGGTC 240  
QY 241 TTCTTCAACGAGAGAAAGTGGGAGAGAAATCACAATTGAATACCTTCCCTTCAAGTAGAT 300  
DB 241 TTCTTCAACGAGAGAAAGTGGGAGAGAAATCACAATTGAATACCTTCCCTTCAAGTAGAT 300  
QY 301 GTGACGGGGGAAAGTGAATCCCGAGAGAAACGAACTCAGGGTGGTTGTGAGAACAGATTG 360  
DB 301 GTGACGGGGGAAAGTGAATCCCGAGAGAAACGAACTCAGGGTGGTTGTGAGAACAGATTG 360  
QY 361 AAAGTGGGAGGATTTCCCTCGAAGGTTCCAGACAGCGGCACTCACACCGTGGGATTTT 420  
DB 361 AAAGTGGGAGGATTTCCCTCGAAGGTTCCAGACAGCGGCACTCACACCGTGGGATTTT 420  
QY 421 GGAAGTTTCCACCTGCAAACTTCACTTCTTCCCTTCCGTTGGAATCATAGGCTGT 480  
DB 421 GGAAGTTTCCACCTGCAAACTTCACTTCTTCCCTTCCGTTGGAATCATAGGCTGT 480  
QY 481 CTGATAGAGTTCAACAGACCGGAGGATCTCGACATCTGGGTGACACGAGTGAAGTCT 540  
DB 481 CTGATAGAGTTCAACAGACCGGAGGATCTCGACATCTGGGTGACACGAGTGAAGTCT 540  
QY 541 GAACCGGAGAGAACTTGAAGAAAGTGAAGTGAAGTGAAGTCTCAGAGAAAGCGGTG 600  
DB 541 GAACCGGAGAGAACTTGAAGAAAGTGAAGTGAAGTGAAGTCTCAGAGAAAGCGGTG 600  
QY 601 GGACAGAGAGATGACGATCAAACTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660  
DB 601 GGACAGAGAGATGACGATCAAACTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660  
QY 661 TTCGTGAAAGGGAGTTCACTCTCGAAAACGCCAGGTTCTGAGACCTCGAAGATCCATAT 720  
DB 661 TTCGTGAAAGGGAGTTCACTCTCGAAAACGCCAGGTTCTGAGACCTCGAAGATCCATAT 720  
QY 721 CTTTATCTCTCAAGGTGGAATTTGAAAAGAGAGAGTACACTCTGAGACATCGGAATCAGA 780  
DB 721 CTTTATCTCTCAAGGTGGAATTTGAAAAGAGAGAGTACACTCTGAGACATCGGAATCAGA 780  
QY 781 ACGATCAGCTGGGAGAGAGAGGCTCTAATCTGAACGGGAAACCTGCTTTTGAAGGCG 840  
DB 781 ACGATCAGCTGGGAGAGAGAGGCTCTAATCTGAACGGGAAACCTGCTTTTGAAGGCG 840  
QY 841 TTTGAAAGCAGAGGAATCCCGGTTCTGGGGCAGGGCACTTTTATTCATGTATGATA 900  
DB 841 TTTGAAAGCAGAGGAATCCCGGTTCTGGGGCAGGGCACTTTTATTCATGTATGATA 900  
QY 901 AAAGACTTCAACCTTCTGAAGTGAATCAACGGAATTTCTTCAAGACCTCTCACTATCT 960  
DB 901 AAAGACTTCAACCTTCTGAAGTGAATCAACGGAATTTCTTCAAGACCTCTCACTATCT 960  
QY 961 TACAGTGAAGAGTGGTGAATCTTCCGACAGACTCGGAATCTTGTGATAGACGAAGCC 1020  
DB 961 TACAGTGAAGAGTGGTGAATCTTCCGACAGACTCGGAATCTTGTGATAGACGAAGCC 1020  
QY 1021 CCGCACGTTGCTATCAAGGTACCACTAATCCCGAGACTCAGAGATAGACAGAGAC 1080  
DB 1021 CCGCACGTTGCTATCAAGGTACCACTAATCCCGAGACTCAGAGATAGACAGAGAC 1080  
QY 1081 AACATTAAGAAATGATGACAGACAGAGAAACCATCCAGTGTGATCATGTGAGGTG 1140  
DB 1081 AACATTAAGAAATGATGACAGACAGAGAAACCATCCAGTGTGATCATGTGAGGTG 1140  
QY 1141 GCGAAGCAACAGAGTCCAACCATCAGACGCGAGGGTTTCTTCAAGCCCTTATGAG 1200

DB 1141 GCGAAGCAACAGAGTCCAACCATCAGACGCGAGGGTTTCTTCAAGCCCTTATGAG 1200  
QY 1201 ACTGCCAATGAATGATGACGAACACGCGCTTGTTCATGTGAGCATGATGAGCGACCA 1260  
DB 1201 ACTGCCAATGAATGATGACGAACACGCGCTTGTTCATGTGAGCATGATGAGCGACCA 1260  
QY 1261 GACGAGAGAACAGAGAGCGTGGCGCTGAAGTACTTCGACATCGTCTGTGTGAACAGTAC 1320  
DB 1261 GACGAGAGAACAGAGAGCGTGGCGCTGAAGTACTTCGACATCGTCTGTGTGAACAGTAC 1320  
QY 1321 TACGCTGTGATCATCTATTCAGGGAAGGATAGAGAGGACTTCAAGCTCTGAAAAAGAC 1380  
DB 1321 TACGCTGTGATCATCTATTCAGGGAAGGATAGAGAGGACTTCAAGCTCTGAAAAAGAC 1380  
QY 1381 ATAGAGAGCTCTATGCAAGGACACAGAAAGCCCATCTTGTGCACAGAAATTGCGTGGGAC 1440  
DB 1381 ATAGAGAGCTCTATGCAAGGACACAGAAAGCCCATCTTGTGCACAGAAATTGCGTGGGAC 1440  
QY 1441 GCGATAGCTGGCATCACTACGATCCACTCAATGTCTCCGAAAGTACCAAGCAGAG 1500  
DB 1441 GCGATAGCTGGCATCACTACGATCCACTCAATGTCTCCGAAAGTACCAAGCAGAG 1500  
QY 1501 CTCGTTGAAGAAAGACGATCAGGCTCCTTTTGAAGAAAGACTACATCATCGGAACACG 1560  
DB 1501 CTCGTTGAAGAAAGACGATCAGGCTCCTTTTGAAGAAAGACTACATCATCGGAACACG 1560  
QY 1561 TGGGCTTTGACAGATTTTAAGATCTCTCAGATGTGAGAGAGCCATTTCAACCAAG 1620  
DB 1561 TGGGCTTTGACAGATTTTAAGATCTCTCAGATGTGAGAGAGCCATTTCAACCAAG 1620  
QY 1621 GGTGTTTTCACAAGAGACAGACCAACCCAACTCGTTGCTCATGTACTGAGAGACTGTG 1680  
DB 1621 GGTGTTTTCACAAGAGACAGACCAACCCAACTCGTTGCTCATGTACTGAGAGACTGTG 1680  
QY 1681 AGTGAGTT 1689  
DB 1681 AGTGAGTT 1689

RESULT 2  
US-10-369-493-26682  
; Sequence 26682, Application US/10369493  
; Publication No. US2003023675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; PRIOR FILING DATE: 2003-02-28  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 26682  
; LENGTH: 1692  
; TYPE: DNA  
; ORGANISM: Thermotoga maritima  
US-10-369-493-26682

Query Match 99.9%; Score 1688; DB 15; Length 1692;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1688; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 ATGTTAAGACCGCAAGAAACAAGAGATTATCTTAATCTTGAATGAGTTGAAAT 60  
DB 1 ATGTTAAGACCGCAAGAAACAAGAGATTATCTTAATCTTGAATGAGTTGAAAT 60  
QY 61 CTTGAAGTAACGAGAAAGACAGACCAATCGCCGTTCTGGAAGCTGGAATGACAGTAC 120

Db 61 CTTGAAGTAACGACAAAGACAGACCAATCGCCGTTCTGAGCTGGAATGAGACGTAC 120  
QY 121 CAGGATCTGTGCTACGAAGAAGACCCCTTCACTCAAAAACCACTTCTACGTTCCGAAG 180  
Db 121 CAGGATCTGTGCTACGAAGAAGACCCCTTCACTCAAAAACCACTTCTACGTTCCGAAG 180  
QY 181 NAACTTTCACAAAAACACATCAGACTTTACTTTGCTGCGGTGAACACGGACTGCGAGTC 240  
Db 181 GAACTTTCACAAAAACACATCAGACTTTACTTTGCTGCGGTGAACACGGACTGCGAGTC 240  
QY 241 TTCTCAACGAGAGAAAAGTGGGAGAGATCACTATTGAATACCTTCCCTTGAAGTAGAT 300  
Db 241 TTCTCAACGAGAGAAAAGTGGGAGAGATCACTATTGAATACCTTCCCTTGAAGTAGAT 300  
QY 301 GTGACGGGGAAAAGTGAATCCCGAGAGAACGAATCAGGGTGTGTGAGAACAGATTG 360  
Db 301 GTGACGGGGAAAAGTGAATCCCGAGAGAACGAATCAGGGTGTGTGAGAACAGATTG 360  
QY 361 AAGTGGGAGGATTTCCCTCGAAGTTCCAGACAGCGGCACTCACACCGTGGATTTTT 420  
Db 361 AAGTGGGAGGATTTCCCTCGAAGTTCCAGACAGCGGCACTCACACCGTGGATTTTT 420  
QY 421 GGAAGTTTTCACCTGCAAACTTGAATTCTTCCCTACGGTGAATCATAGGCTGTT 480  
Db 421 GGAAGTTTTCACCTGCAAACTTGAATTCTTCCCTACGGTGAATCATAGGCTGTT 480  
QY 481 CTGATAGAGTTACAGACCACGCGAGGATACTCGACATCTGGTGGACACGAGTGAAGTCT 540  
Db 481 CTGATAGAGTTACAGACCACGCGAGGATACTCGACATCTGGTGGACACGAGTGAAGTCT 540  
QY 541 GAAACCGGAGAGAAAATTGGAAAAGTGAAGTGAAGTGAAGTCTCAGAAAGACGGTG 600  
Db 541 GAAACCGGAGAGAAAATTGGAAAAGTGAAGTGAAGTGAAGTCTCAGAAAGACGGTG 600  
QY 601 GGAACGAGAGATGACGATCAAACTTGAAGAGAGAGAAAAGATTAGAACATCCACAGA 660  
Db 601 GGAACGAGAGATGACGATCAAACTTGAAGAGAGAGAAAAGATTAGAACATCCACAGA 660  
QY 661 TTCGTCGAAGGGGAGTTTCATCTCGAAAAACGCGAGTTCTGAGCTCGAAGATCCATAT 720  
Db 661 TTCGTCGAAGGGGAGTTTCATCTCGAAAAACGCGAGTTCTGAGCTCGAAGATCCATAT 720  
QY 721 CTTTATCTCTCAAGGTGGAACCTGAAAAAGACGAGTACACTCTGGAATCGGAATCAGA 780  
Db 721 CTTTATCTCTCAAGGTGGAACCTGAAAAAGACGAGTACACTCTGGAATCGGAATCAGA 780  
QY 781 ACGATCAGCTGGGACGAGAGAGGCTTATCTGAACGGAAAACTGTCTTTTGAAGGCG 840  
Db 781 ACGATCAGCTGGGACGAGAGAGGCTTATCTGAACGGAAAACTGTCTTTTGAAGGCG 840  
QY 841 TTTGGAAGCAGAGGAATTCGCCGTTCTGGGGCAGGGCACCTTTATCCATTGATGATA 900  
Db 841 TTTGGAAGCAGAGGAATTCGCCGTTCTGGGGCAGGGCACCTTTATCCATTGATGATA 900  
QY 901 AAAGACTTCAACCTTCTGAAGTGAATCAACGGAATTTCTTACAGACTCTCACTATCCT 960  
Db 901 AAAGACTTCAACCTTCTGAAGTGAATCAACGGAATTTCTTACAGACTCTCACTATCCT 960  
QY 961 TACAGTGAAGAGTGGCTGGATCTTGCCGACAGACTCGGAATCTTGTGATAGAGAAAGCC 1020  
Db 961 TACAGTGAAGAGTGGCTGGATCTTGCCGACAGACTCGGAATCTTGTGATAGAGAAAGCC 1020  
QY 1021 CCGCAGCTGGTATCAACAAGGTACCACTAACCCGAGACTCAGAAAGATAGCAGAAAGAC 1080  
Db 1021 CCGCAGCTGGTATCAACAAGGTACCACTAACCCGAGACTCAGAAAGATAGCAGAAAGAC 1080  
QY 1081 AACATTAAGAGAATGATCGACAGACACAAGAACCAATCCAGTGTGATCATGTGAGTGTG 1140  
Db 1081 AACATTAAGAGAATGATCGACAGACACAAGAACCAATCCAGTGTGATCATGTGAGTGTG 1140  
QY 1141 GCGAAGCAACCAAGAGTCCAAACCATCCAGACGCGGAGGGTTCTTCAAAAGCCCTTATAG 1200

Db 1141 GCGAAGCAACCAAGAGTCCAAACCATCCAGACGCGGAGGGTTCTTCAAAAGCCCTTATAG 1200  
QY 1201 ACTGCCAATGAATGATGATGAAACACGCCCGTTGTATGTGATGATGATGAGCCACCA 1260  
Db 1201 ACTGCCAATGAATGATGATGAAACACGCCCGTTGTATGTGATGATGATGAGCCACCA 1260  
QY 1261 GACGAGAGAACAGAGACGTTGGGCTGAAGTACTTGCACATCGTGTGTGAACAGGTAC 1320  
Db 1261 GACGAGAGAACAGAGACGTTGGGCTGAAGTACTTGCACATCGTGTGTGAACAGGTAC 1320  
QY 1321 TACGCTGTGATCACTTATCAGGGAAGATAGAGAAGACTTCAAGCTCTGAAAAAGAC 1380  
Db 1321 TACGCTGTGATCACTTATCAGGGAAGATAGAGAAGACTTCAAGCTCTGAAAAAGAC 1380  
QY 1381 ATAGAAGAGCTTATGCAAGGCAAGAAAGCCATCTTGTCAAGAAATTCGTCGGGAC 1440  
Db 1381 ATAGAAGAGCTTATGCAAGGCAAGAAAGCCATCTTGTCAAGAAATTCGTCGGGAC 1440  
QY 1441 GCGATAGCTGGCATCCACTACGATCCACCTCAATGTTCTCCGAAAGATACCAAGCAGAG 1500  
Db 1441 GCGATAGCTGGCATCCACTACGATCCACCTCAATGTTCTCCGAAAGATACCAAGCAGAG 1500  
QY 1501 CTCGTTGAAAAAGAGATCAGGCTCTTTTGAAAAAAGACTACATCGGAACACAGCTG 1560  
Db 1501 CTCGTTGAAAAAGAGATCAGGCTCTTTTGAAAAAAGACTACATCGGAACACAGCTG 1560  
QY 1561 TGGGCTTTTGAGATTTTAAGACTCTCAGAAATGTGAGAAGACCATTTCTAACACAAG 1620  
Db 1561 TGGGCTTTTGAGATTTTAAGACTCTCAGAAATGTGAGAAGACCATTTCTAACACAAG 1620  
QY 1621 GGTGTTTTCACAAGAGACAGACACCAAACTGCTGCTCATGTAAGTGAAGAGCTGTG 1680  
Db 1621 GGTGTTTTCACAAGAGACAGACACCAAACTGCTGCTCATGTAAGTGAAGAGCTGTG 1680  
QY 1681 AGTGAAGTT 1689  
Db 1681 AGTGAAGTT 1689

RESULT 3  
US-10-136-841-5  
; Sequence 5, Application US/10136841  
; Publication No. US20030082176A1  
; GENERAL INFORMATION:  
; APPLICANT: Lebowitz, Jonathan  
; APPLICANT: Beverly, Stephen  
; TITLE OF INVENTION: SUBCELLULAR TARGETING OF THERAPEUTIC PROTEINS  
; FILE REFERENCE: SYM-007  
; CURRENT APPLICATION NUMBER: US/10/136,841  
; PRIOR FILING DATE: 2002-08-22  
; PRIOR APPLICATION NUMBER: US 60/287,531  
; PRIOR FILING DATE: 2001-04-30  
; PRIOR APPLICATION NUMBER: US 60/304,609  
; PRIOR FILING DATE: 2001-07-10  
; PRIOR APPLICATION NUMBER: US 60/329,461  
; PRIOR FILING DATE: 2001-10-15  
; PRIOR APPLICATION NUMBER: US 60/351,276  
; PRIOR FILING DATE: 2002-01-23  
; NUMBER OF SEQ ID NOS: 22  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 5  
; LENGTH: 2169  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: A recombinant sequence incorporating a signal peptide sequence,  
; OTHER INFORMATION: the mature human beta-glucuronidase sequence, a bridge of three  
; OTHER INFORMATION: amino acids, and an IGF-II sequence  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(2166)  
; US-10-136-841-5



Query Match 10.3%; Score 174; DB 14; Length 2169;  
 Best Local Similarity 52.2%; Pred. No. 1.2e-44;  
 Matches 493; Conservative 0; Mismatches 430; Indels 21; Gaps 4;

```

QY 737 TGAACCTGAAAAAGAGAGTACCTCTGACATCGAATCAGATCAGCTGGAGC 796
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 941 TGGGGCTGTGTCTGACTTCTACACACTCCCTGGGGATCCGCACTGTGCTGCACCA 1000

QY 797 AGAAGAGCTCTATCTGAACGGGAAACCTGTCTTTTGAAGGGCTTTGAAAGCAGAG 856
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1001 AGAGCCAGTTCCTCATTAATGGGAAACCTTTCTATTTCACGGGTGCAACAAGATGAGG 1060

QY 857 AATTCCTCGTCTGGGGCAGGGCAGCTTTTATCATTGATGATAAAGACTTCAACCTTC 916
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1061 ATGGGACATCCGAGGGGAGGGCTTGAAGCTGGCCCTGCTGTGTAAGGACTTCAACCTGC 1120

QY 917 TGAAGTGGATCAACGGCAATCTTTTCAGACCTCTCACTATCTTACAGTGAAGAGTGC 976
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1121 TTGCTGGCTTGGTCCCAACGCTTTCCTACCACTACCCCTATGACAGAGAACTGA 1180

QY 977 TGGATCTTGGCCGACAGACTCGGAATCTGTGATAGACGAAGCCCGCAGCTGGTATCA 1036
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1181 TGCAGATGTGTGACCGCTATGGGATTTGTGTATGATGATGTCTCCGGCGTGGCTGTGG 1240

QY 1037 CAAGGTACCA---CTACAATCCCGAGACTCAGAAAGATAGCAGAAACACATTAAGAGAA 1093
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1241 CGCTGCCGAGTCTTCAACAACGTTTCTCTGATCACCACATGACAGGTGATGGAAGAA 1300

QY 1094 TGAATGACAGACACAAGAACCATCCAGTGTGATCATGTGAGTGTGGCGAAAGAACAG 1153
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1301 TGGTGGCGTAGGACACAAGAACCCCGCGGTGATGTGTGTGTGTGTGTGTGTGTGTGTGT 1360

QY 1154 AGTCCAACCATCCAGACGCGGAGGTTTCTTCAAGCCCTTTATGAGACTGGCAATGAAA 1213
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1361 CGTCCCACTAGAAATCTGTGCTACTACTTGAAGATGTGTGTGTGTGTGTGTGTGTGTGT 1420

QY 1214 TGGATCGAACACGCCCCGTTGTGATGTGAGACATGATGAGCAGCAGCAGAGAGAAACAA 1273
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1421 TGGACCCCTCCCGGCTGTGACTTTGTGAGCACTTAATATGAG----- 1468

QY 1274 GAGACGTGGCGCTGAAGTACTTCGACATCGTGTGTGAAAGAGTACTAGCGCTGTACA 1333
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1469 CAGACAAAGGGGCTCCGTATGTGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1528

QY 1334 TCTATCAGGAAAGATAGAGAAAGACTTCAAGCTCTGAAAGAAAGACATAGAGAGCTCT 1393
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1529 ACGACTACGGGCACTGAGTGTGATTCAGTGCAGCTGCGCCACCCAGTTGAGAACTGGT 1588

QY 1394 ATGCAAGGACAGAAAGCCCATCTTGTGACAGAAATTCGGTGGCGGACCGCATAGCTGCA 1453
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1589 ATTAAGAGATTCAGAAAGCCCATTTATTCAGAGCAGATATGAGCAGAAACGATTTGAGGG 1648

QY 1454 TCCACTACGATCCACCTCAAAATGTTCTCCGAAAGATACCAAGCAGAGCTCGTTGAAAG- 1512
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1649 TTCACAGAGATCCACCTCTGATGTTCATGAAAGATACCAAGAAAGTCTGCTAGAGCAGT 1708

QY 1513 ACGATCAGGCTCTTTTGAAGAA-----AGACTACATCATCGGAACACACGCTGGCGCT 1567
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1709 ACCATCTGGGTCTGATCAAAAGCAGAAATATGTGTGTGAGAGCTCATTTGGAATTT 1768

QY 1568 TTGAGATTTTAAAGTCTCTCAGATGTGAGAGAGACCCATTTCTCAACACAGAGGTGTTT 1627
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1769 TTGCGGATTTTATGACTGAACAGTCAACGACGAGAGTGTGGGGAATTAAGGGGATCT 1828

QY 1628 TCACAAAGAGACAGACAAACCAACTGTTGCTCATGTACTGAGA 1671
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1829 TCACTCGGACAGACAAACCAAAAGTGCAGCGTTCTTTTGGCA 1872
  
```

## RESULT 4

US-10-272-531A-5  
 ; Sequence 5, Application US/10272531A  
 ; Publication No. US20040005309A1

```

; GENERAL INFORMATION:
; APPLICANT: Lebowitz, Jonathan H
; APPLICANT: Beverly, Stephen
; APPLICANT: SLY, William S.
; TITLE OF INVENTION: TARGETED THERAPEUTIC PROTEINS
; FILE REFERENCE: SYM-009
; CURRENT APPLICATION NUMBER: US/10/272,531A
; PRIORITY FILING DATE: 2002-10-16
; PRIORITY APPLICATION NUMBER: US 60/384,452
; PRIORITY FILING DATE: 2002-05-29
; PRIORITY APPLICATION NUMBER: US 60/386,019
; PRIORITY FILING DATE: 2002-06-05
; PRIORITY APPLICATION NUMBER: US 60/408,816
; PRIORITY FILING DATE: 2002-09-06
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 2169
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: A recombinant DNA sequence incorporating a signal peptide sequenc
; OTHER INFORMATION: e, the mature human beta-glucuronidase sequence, a bridge of thre
; OTHER INFORMATION: e amino acids, and an IGF-II sequence
; NAME/KEY: CDS
; LOCATION: (1)..(2166)
; OTHER INFORMATION:
; US-10-272-531A-5
  
```

Query Match 10.3%; Score 174; DB 16; Length 2169;  
 Best Local Similarity 52.2%; Pred. No. 1.2e-44;  
 Matches 493; Conservative 0; Mismatches 430; Indels 21; Gaps 4;

```

QY 737 TGAACCTGAAAAAGAGAGTACCTCTGACATCGAATCAGATCAGCTGGAGC 796
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 941 TGGGGCTGTGTCTGACTTCTACACACTCCCTGGGGATCCGCACTGTGCTGCACCA 1000

QY 797 AGAAGAGCTCTATCTGAACGGGAAACCTGTCTTTTGAAGGGCTTTGAAAGCAGAG 856
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1001 AGAGCCAGTTCCTCATCAATGGGAAACCTTTCTATTTCACCGGTGCAACAAGATGAGG 1060

QY 857 AATTCCTCGTCTGGGGCAGGGCAGCTTTTATCCATTGATGATAAAGACTTCAACCTTC 916
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1061 ATGGGACATCCGAGGGGAGGGCTTTCAGTGGCCGCTGTGTGGAAGAGACTTCAACCTGC 1120

QY 917 TGAAGTGGATCAACGGCAATCTTTCAGACCTCTCACTATCTTACAGTGAAGAGTGC 976
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1121 TTGCTGGCTTGGTCCCAACGCTTTCCTGACAGCAGTACCTTATGACAGAGAACTGA 1180

QY 977 TGGATCTTGGCCGACAGACTCGGAATCTTGTGATAGAGAAAGCCCGCAGCTGGTATCA 1036
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1181 TGCAGATGTGTGACCGCTATGGGATTTGTGTGATGATGATGTGTGTGTGTGTGTGTGTGTGT 1240

QY 1037 CAAGGTACCA---CTACAATCCCGAGACTCAGAAAGATAGCAGAAAGACAAATTAAGAGAA 1093
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1241 CGTCCGCACTTCTTCAACAACGTTTCTGTGATCACCACATGACAGTGTGGAAGAAAG 1300

QY 1094 TGAATGACAGACACAAGAACCATCCAGTGTGATCATGTGAGTGTGGCGAAAGAACAG 1153
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1301 TGGTGGTAGGACACAAGAACCCCGCGGTGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1360

QY 1154 AGTCCAACCATCCAGACGCGGAGGTTTCTTCAAGCCCTTATGAGACTGCCAATGAAA 1213
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1361 CGTCCCACTAGAAATCTGTGCTACTACTTGAAGATGTGTGTGTGTGTGTGTGTGTGTGTGT 1420

QY 1214 TGGATCGAACACGCCCCGTTGTGATGTGAGACATGATGAGCAGCAGCAGAGAGAGAAACAA 1273
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1421 TGGACCCCTCCCGGCTGTGACTTTGTGAGCACTTAATATGAG----- 1468

QY 1274 GAGACGTGGCGCTGAAGTACTTCGACATCGTGTGTGAAACAGTACTAGCGCTGTACA 1333
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1469 CAGACAAAGGGGCTCCGTATGTGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1528
  
```



APPLICANT: Yang, Bin  
TITLE OF INVENTION: RECOMBINANT ADENO-ASSOCIATED VIRUS VIRIONS FOR THE  
FILE OF INVENTION: TREATMENT OF LYSOSOMAL DISORDERS  
FILE REFERENCE: 0800-0021  
CURRENT APPLICATION NUMBER: US/10/421,175  
CURRENT FILING DATE: 2003-04-22  
PRIOR APPLICATION NUMBER: US/09/715,858  
PRIOR FILING DATE: 2000-11-14  
NUMBER OF SEQ ID NOS: 16  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 1  
LENGTH: 1956  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1)..(1956)  
US-10-421-175-1

Query Match 10.2%; Score 172.4; DB 15; Length 1956;  
Best Local Similarity 52.1%; Pred. No. 3.6e-44;  
Matches 492; Conservative 0; Mismatches 431; Indels 21; Gaps 4;

QY 737 TGGAACTTGAAGAAAGAGTACCTCTGACATCGAATCGAATGAAAGATGATGAGTGGAGC 796  
DB 938 TGGGGCTGTGTCTGACTTCTACACACTCCCTGTGGGATCCGCACTGTGCTGCACCA 997  
QY 797 AGAAGAGGCTCTATCTGAACGGGAACTGTCTTTTGAAGGCTTTGAAAGCAGAGG 856  
DB 998 AGAGCCAGTCTCTATCATCAATGGGAACTTTCTATTTCCACGGTGTCAACAAGCATGAGG 1057  
QY 857 AATTCCCGTCTGGGGGAGGCACTTTTATCCATTGATGATTAAGAACTTCAACCTTC 916  
DB 1058 ATGGGACATCCGAGGAGGAGGCTTGAAGTGGCTGCTGTGGAAGATTCACCTGC 1117  
QY 917 TGAAGTGATCAACGGAATCTTTCAGAGCTCTCACTATCTTACAGTGAAGTGGC 976  
DB 1118 TTGCTGCTGTGGTGGCAACGCTTTCCTGACCACTACCTCCCTATGAGAGAAAGTGA 1177  
QY 977 TGGATCTTGGCGACAGACTCGGAATCTTGTATGAGAGAGCCCGCAGCTTGTATCA 1036  
DB 1178 TGCAGATGTGTACCGCTATGGAGTGTGTGATCATGATGATGTCCCGGCTGGCTGG 1237  
QY 1037 CAAGGTACCA---CTACAATCCCGAGACTCGAAGATAGCAGAAAGACAAATAGAAGA 1093  
DB 1238 CGCTGCCGCACTTCTTCAACAACGTTTCTTGTGATCAACCATGAGTGAAGAAAG 1297  
QY 1094 TGATGACAGACAAAGAACATCCAGTGTGATCATGTGAGTGGGAAAGAACAG 1153  
DB 1298 TGGTGGTGGAGCAAGAACACCCCGGCTGTGATGTGTGTGCTGTGCTCAAGAGCTG 1357  
QY 1154 AGTCCAAACCATCCAGACGCGGAGGCTTCTTCAAGCCCTTATGAGACTGCCAATGAA 1213  
DB 1358 CGTCCCACTAGAAATCTGTGCTGCTACTACTGAAGATGTGATCGCTCAACCAATCT 1417  
QY 1214 TGGATCGAAGACGCCCCCTGTGTGATGTGAGCATGTGAGCGACCAAGAGAGAA 1273  
DB 1418 TGGACCCCTCCCGCTGTGACCTTGTGAGCAACTTAATATGAG----- 1465  
QY 1274 GAGACGTGGCGCTGAAGTACTTGCACATCGTGTGTGAGACAGTACTACGGCTGTACA 1333  
DB 1466 CAGACAAAGGGGCTCCGTAATGTGATGTGATGTGTAACAGCTAATCTTGTGATTC 1525  
QY 1334 TCTATCAGGGAAGATAGAGAAAGACTTCAAGCTCTGGAAGAAAGACATAGAAGCTCT 1393  
DB 1526 ACGACTACGGGCACTGGAGTGTATTACGTGACGTGCGCACCAGTTTGAAGACTGTG 1585  
QY 1394 ATGCAAGGACAGAAAGCCATCTTTGTGACAGAAATCGGTGCGGACGATAGCTGGCA 1453  
DB 1586 ATAGAGATATCAGAAAGCCATTAATCAGAGCGAGTATGAGACAGAAAGATTCAGGGT 1645  
QY 1454 TCCACTACGATCCACCTCAAAATGTTCTCGAAGAGTACCAAGCAGAGCTGTTGAAAG- 1512  
DB 1512 TGGTGGTGGAGCAAGAACCAACCCCGGCTGTGATGTGTGTGCTGTGCTGCAAGAGCTG 1383

DB 1646 TTCACCAGATCCACCTGTGATGTCTACTGAGAGTACCAGAAAGTCTGTAGACACT 1705  
QY 1513 ACGATCAGGCTCCTTTGAAAAA-----AGACTACATCATCGAAGACACAGTGGGCT 1567  
DB 1706 ACCATCTGGGTCTGATCAAAAAACGAGAAATATGTGTGAGAGCTCATTTGGAAT 1765  
QY 1568 TTGCAATTTTAAGCTCCTCAGAAATGTGAGAGAACCAATTTCAACCAAGAGGTTT 1627  
DB 1766 TTGCCGATTTTATGATGATCAAGAGTCAACCGACGAGAGCTGGGAAATAAAAAGGATCT 1825  
QY 1628 TCACAAGAGACAGACAAACCAACTCTGCTCATGTACTGAGA 1671  
DB 1826 TCACCTGGCAGAGACAAACCAAAAGTGCAGCGTTCTTTGCGA 1869

RESULT 7  
US-10-388-360-343

Sequence 343, Application US/10388360  
Publication No. US20030225528A1  
GENERAL INFORMATION:  
APPLICANT: GENOMIC HEALTH  
APPLICANT: Baker, Joffe B.  
APPLICANT: Cronin, Maureen T.  
APPLICANT: Kiefer, Michael C.  
APPLICANT: Shak, Steve  
APPLICANT: Walker, Michael Graham  
TITLE OF INVENTION: GENE EXPRESSION PROFILING IN BIOPSED TUMOR TISSUES  
FILE REFERENCE: 39740-0001US  
CURRENT FILING DATE: 2003-03-12  
PRIOR APPLICATION NUMBER: US 60/412,049  
PRIOR FILING DATE: 2002-09-18  
PRIOR APPLICATION NUMBER: US 60/364,890  
PRIOR FILING DATE: 2002-03-13  
NUMBER OF SEQ ID NOS: 384  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 343  
LENGTH: 2191  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-10-388-360-343

Query Match 10.2%; Score 172.4; DB 15; Length 2191;  
Best Local Similarity 52.1%; Pred. No. 3.6e-44;  
Matches 492; Conservative 0; Mismatches 431; Indels 21; Gaps 4;

QY 737 TGGAACTTGAAGAAAGAGTACCTCTGACATCGAATCGAATGAAAGATGATGAGTGGAGC 796  
DB 964 TGGGGCTGTGTCTGACTTCTACACACTCCCTGTGGGATCCGCACTGTGCTGCACCA 1023  
QY 797 AGAAGAGGCTCTATCTGAACGGGAACTGTCTTTTGAAGGCTTTGAAAGCAGAGG 856  
DB 1024 AGAGCCAGTCTCTATCATCAATGGGAAACCTTTCTATTTCCACGGTGTCAACAAGCATGAG 1083  
QY 857 AATTCCCGTCTGGGGGAGGCACTTTTATCCATTGATGATTAAGAACTTCAACCTTC 916  
DB 1084 ATGGGACATCCGAGGAGGAGGCTTGCAGTGGCCGCTGTGTAAGAGACTTCAACCTGC 1143  
QY 917 TGAAGTGATCAACGGAATCTTTCAGAGCTCTCACTATCTTACAGTGAAGAGTGGC 976  
DB 1144 TTGCTGCTGTGGTGGCAACGCTTTCCTGACCAAGCACTTATGAGAGAAAGTGA 1203  
QY 977 TGGATCTTGGCGACAGACTCGGAATCTTGTATGAGAGAGCCCGCAGCTTGTATCA 1036  
DB 1204 TGCAGATGTGTACCGCTATGGAGTGTGTGATCATGATGATGTCCCGGCTGGCTGG 1263  
QY 1037 CAAGGTACCA---CTACAATCCCGAGACTCGAAGATAGCAGAAAGACAAATAGAAGAA 1093  
DB 1264 CGCTGCCGCACTTCTTCAACAAGCTTCTGTGATCAACCATGAGTGAAGAG 1323  
QY 1094 TGATGACAGACAAAGAACATCCAGTGTGATCATGTGAGTGGGAAAGAACAG 1153  
DB 1324 TGGTGGTGGAGCAAGAACCAACCCCGGCTGTGATGTGTGTGCTGTGCTGCAAGAGCTG 1383

[illegible]

## RESULT 8

```

US-10-429-802-16
; Sequence 16, Application US/10429802
; Publication No. US20030228285A1
; GENERAL INFORMATION:
; APPLICANT: HUNG, MIEN-CHIE
; APPLICANT: WONG, KA YIN
; APPLICANT: ZOU, YIYU
; TITLE OF INVENTION: BIPARTITE T-CELL FACTOR (TCF)-RESPONSIVE PROMOTER
; FILE REFERENCE: UTSC:752US
; CURRENT APPLICATION NUMBER: US/10/429,802
; CURRENT FILING DATE: 2003-05-05
; PRIOR APPLICATION NUMBER: 60/377,672
; PRIOR FILING DATE: 2002-05-03
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 16
; LENGTH: 2191
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-429-802-16

```

Query Match	10.2%;	Score 172.4;	DB 15;	Length 2191;
Best Local Similarity	52.1%;	Pred. No. 3.8e-44;		
Matches 492;	Conservative 0;	Mismatches 431;	Indels 21;	Gaps 4;
QY	737	TGGAACCTGAAAAAGACGAGTACACTCTGGACATCGGAATCAGAACGATCAGCTGGAGC	796	
Db	964	TGGGGCCTGTGTCTGACTTCTACACACTCCCTGTGGGATCCGCACTGTGGCTGTACCA	1023	
QY	797	AGAAGAGGCTCTATCTGAAACGGGAAACCTGTCTTTTGAAGGGCTTTGGAAGCAGGAGG	856	
Db	1024	AGAGCCAGTTCCTCATCAATGGGAAACCTTTCTATTTCCACGGGTGCAACAAGCATGAGG	1083	
QY	857	AATTCCTCCGTTCTGGGGCAGGGAACCTTTATTCATTGATGATAAAGACTTCAACCTTC	916	

Db	1084	ATGCGACATCCGAGGGAAGGCGCTTCGACTGGCCGCTGCTGGTGAAGACTTCAACCTGC	1143
Qy	917	TGAAGTGATCAACGGAAATCTTCAGAGCCTCTCACTATCCTTACAGTGAAGAGTGGC	976
Db	1144	TTCGCTGGCTTGGTGCCAAACGCTTCCGTACAGCCACTACCCCTATGCAGAGAAGTGA	1203
Qy	977	TGGATCTTGGCGACAGACTCGGAATCCTGTGTATAGACGAAGCCCGCAGCTGTGTATCA	1036
Db	1204	TGCAGATGTGTACCGCTATGGGATGTGTGCATGCATGATGTGCCGGGCTGGCCTGG	1263
Qy	1037	CAAGGTACCA---CTACAATCCGAGACTCAGAAGATAGCAGAAGACAATAGAAGA	1093
Db	1264	CGGTGCCGAGTTCTTCAACAAGTTCTCTGCATCACCAATGCAGGTGATGAAGAAG	1323
Qy	1094	TGATCGACAGACAGAAGAACATCCAGTGTGATCATGTGGAAGTGTGGCAGAACAG	1153
Db	1324	TGGTGCGTAGGAGACAAGAACACCCCGCGCTGTGATGTGTCTGTGGCCAAGAGCCTG	1383
Qy	1154	AGTCCAACCATCCAGACGCGGAGGGTTCCTTCAAGCCCTTATGAGACTGCCAATGAA	1213
Db	1384	CGTCCACCTAGAAATCTGCTGCTACTACTTGAAGATGGTGAATCGCTCACACCAATCCT	1443
Qy	1214	TGGATCGAACACGCCCCGTTGTTCATGTTGAGCATGATGGACGCCACGACGAGAACA	1273
Db	1444	TGGAACCCCTCCCGGCTGTGACCTTGTGTGACCACTCTAATATGCAG-----	1491
Qy	1274	GAGACGTGGCGCTGAAGTACTTCGACATCGTCTGTGTGAACAGTACTACGCGCTGTACA	1333
Db	1492	CAGACAAAGGGGCTCCGTATGTGATGTGATCTGTTTGAACAGCTACTCTTGTGTATC	1551
Qy	1334	TCTATCAGGGAAGATTAGAAGAAGACTTCAAGCTCTGGAAGAAAGACATAGAAGACTCT	1393
Db	1552	ACGACTACGGGCACTGTGAGTTGATTACGTGCAGCTGGGCCACCCAGTTTGAAGACTGGT	1611
Qy	1394	ATGCAAGGCACAGAAGGCCCATCTTGTGTACAGAATTCGGTGGGACGGGATAGCTGGCA	1453
Db	1612	ATAAGAAGTATCAGAAGCCCATTTATTCAGAGCGAGTATGAGACAGAAACGATTGCAAGGT	1671
Qy	1454	TCCACTACGATCCACTCAATGTTCTCCGAAGAGTACCAAGCAGAGCTGTGAAAG-	1512
Db	1672	TTTCAACGAGATCCACTCTGTATGTTCACTGAAGAGTACCGAAGAAAGTGTGCTAGAGCAGT	1731
Qy	1513	ACGATCAGGCTCTTTGAAAAA----AGACTACATCATCGGAACACACAGTGTGGGCT	1567
Db	1732	ACCATCTGGGCTGTGATCAAAAAACGAGAAATATGTGGTTGGAGAGCTCATTTGGAATT	1791
Qy	1568	TTGCAGATTTTAAGACTCTTCAGAATGTGGAAGAACCCATTCTCAACCAAGGGGTGTTT	1627
Db	1792	TTGCCGATTTCACTAGTGAACAGTCAACCGACGAGAGTGTGGGGAATAAAAAGGGGATCT	1851
Qy	1628	TCACAAGAGACAGAACCAACCAACTCGTTGCTCATGTACTGAGA 1671	
Db	1852	TCACCTGGCAGAGACAAACCAAAAGTGACGCTTCTTTTGGCA 1895	

## RESULT 9

```

US-10-430-503-7
; Sequence 7, Application US/10430503
; Publication No. US20040005684A1
; GENERAL INFORMATION:
; APPLICANT: HUNG, MIEN-CHIE
; APPLICANT: LAN, KENG-LI
; APPLICANT: OU-YANG, FU
; APPLICANT: LIU, JAW-CHING
; APPLICANT: LAN, KENG-HSIN
; TITLE OF INVENTION: TARGETING PROTEINS TO DELIVER THERAPEUTIC OR DIAGNOSTIC
; TITLE OF INVENTION: REAGENTS
; FILE REFERENCE: UTSC:797US
; CURRENT APPLICATION NUMBER: US/10/430,503
; CURRENT FILING DATE: 2003-05-06
; PRIOR APPLICATION NUMBER: 60/383,063
; PRIOR FILING DATE: 2002-05-06

```





Db 1612 ATAGAAGTATCAGAAAGCCCATTTACAGAGCGAGTATGAGCAGAAACGATTCAGAGGT 1671  
QY 1454 TCCACTACGATCCACCTCAATGTTCTCCGAAGATACCAAGCAGAGCTCGTTGAAAAG- 1512  
Db 1672 TTCACCAAGATCCACCTCTGATGTTCACTGAAGAGTACCAAGAAAGTCTGCTAGACAGT 1731  
QY 1513 AGCATCAGGCTCCTTTTGA AAAA-----AGACTACATCATCGGAACACACGCTGGGCT 1567  
Db 1732 ACCATCTGGGTCTGGATCAAAAACGACAGAAATATGTGTTGGAGAGCTCATTTGGAATT 1791  
QY 1568 TTGCAGATTTTAAGACTCTCAGATGTGAGAAACCCATTCTCAACCAAGGCTTTT 1627  
Db 1792 TTGCCGATTTTCATGACTGAACAGTCCGACGAGAGTGTGGGGAATAAAAAGGGATCT 1851  
QY 1628 TCACAAGAGACAGCAACCCAACTCGTTCATGTACTGAGA 1671  
Db 1852 TCACCTCGGACAGACAAACCAAAAGTCAAGCGTTCCTTTGCCGA 1895

RESULT 11

US-10-388-934-225  
; Sequence 225, Application US/10388934  
; Publication No. US20040005547A1  
; GENERAL INFORMATION:  
; APPLICANT: Boeers, Franziska  
; APPLICANT: Suter-Dick, Laura  
; APPLICANT: Wolf, Detlef  
; TITLE OF INVENTION: BIOMARKERS AND EXPRESSION PROFILES FOR TOXICOLOGY  
; FILE REFERENCE: 21199  
; CURRENT APPLICATION NUMBER: US/10/388,934  
; CURRENT FILING DATE: 2003-03-14  
; PRIOR APPLICATION NUMBER: 02005336.9  
; PRIOR FILING DATE: 2002-03-14  
; PRIOR APPLICATION NUMBER: 02015657.6  
; PRIOR FILING DATE: 2002-07-17  
; NUMBER OF SEQ ID NOS: 862  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 225  
; LENGTH: 2472  
; TYPE: DNA  
; ORGANISM: Rattus norvegicus (No. US20040005547A1way rat)  
US-10-388-934-225

Query Match

Best Local Similarity 9.9%; Score 166.8; DB 16; Length 2472;  
Matches 477; Conservative 0; Mismatches 432; Indels 21; Gaps 3;

QY 751 GACGAGTACACTCTGACATCGAATCAGAACGATCAGCTGGACGAGAGAGCTCTAT 810  
Db 946 GACTTCTATACCTCCCTGTGCGGATTCGAACAGTGGCTGTCAAAAAGACAGATTCCTC 1005  
QY 811 CTGAACGGGAACCTGTCTTTTGAAGGGCTTGGAAAACGACGAGGAATTCCTCGTTCTG 870  
Db 1006 ATAAATGGGAAGCTTTTCTACTTCCAAAGCGCTCAACAGCATGAGATTCAAGTATCCGA 1065  
QY 871 GGGCAGGGACCTTTTATCCATGATGATAAAGACTTCAACCTTCTGAAGTGATCAAC 930  
Db 1066 GGGAGAGGCTTCGACTGGCTCTGCTGATAAAGATTCAACCTCCTCCGTTGGCTCGGG 1125  
QY 931 GCGAATTCCTTCAAGACCTCTCACTATCCTTACAGTGAAGAGTGGCTGGATCTTGGCCGAC 990  
Db 1126 GCAAAATTCCTTCTGTAACGACCACTATCCCTACTCGGAGAGGTACTTCAAGCTCTGTGAC 1185  
QY 991 AGACTCGGAATCCTTGTGATAGACGAAAGCCCGCAGCTTGTGATCAAGGTACCA---C 1047  
Db 1186 CGAATATGGAATGTGTGATCATGATGATGATCCCGGTGTGGGATCTGTGCTGCCAGAGT 1245  
QY 1048 TACAATCCCGAAGTACAGAAAGATGACAGAACATTAAGAGATGATCGACAGACAC 1107  
Db 1246 TTTGGCAACGTGTCTTTCGGACCACTAGAGGTGATGACGAGCTGTGTGCGGAGGAC 1305  
QY 1108 AAGAACCATCCAGTGTGATCATGTGAGTGTGGCGAAGCAAGACGAGTCCAAACATCCA 1167

Db 1306 AAAAATCACCTCGCGGTCTGTGATGTGTCTGTGGCCCAATGAGCTGTCTTCTCTGAAA 1365  
QY 1168 GACGCGGAGGTTTCTTCAAAAGCCCTTTATGAGACTGCCAATGAATGATGCAACAGC 1227  
Db 1366 CTGCGCGATATTAATTCAAGACGCTGATCGCCACACCAAGCCCTGACCCCAACCGT 1425  
QY 1228 CCCGTTGTGATGTGAGCATGATGAGACGACCAAGACGAGAGAAACAAGAGCTGGCGCTG 1287  
Db 1426 CCGTGTACCTTTGTGAGCAATA-----CCAGATATGACGACGACATGGGGGCC 1473  
QY 1288 AAGTACTTCGACATCGTCTGTGTGAACAGGTACTACGCGCTGGTACATCTATCAGGGAAG 1347  
Db 1474 CCGTACGTGACGTGATTTGTGTGAACAGTACTTATCCTGTGATCATGACTACGGGCAT 1533  
QY 1348 ATAGAAGAGACTTCAAGCTCTGAAAAACATAGAAGAGCTTATGCAAGGCACAGA 1407  
Db 1534 CTGAGAGTGATTCAGCTGCAGCTGACTAGCCAGTTTGAAGACTGTATTAAGATGTACAG 1593  
QY 1408 AAGCCATCTTTGTACAGAAATTCGGTGGGACCGGATAGCTGGCATCCACTACGATCCA 1467  
Db 1594 AAGCCAATTTATCCAGAGCGAGTATGAGCAGACGCCGCTCTCGGGGCTTCAATGAGATCCA 1653  
QY 1468 CCTCAATGTTCTCCGAAGAGTACCAAGCAGAGCTCGTTGAAAAAGCATCAGGCTCCTT 1527  
Db 1654 CCTGCATGTTCAGTGAGAGTACCAAGACAGCTCTCTGGAGAAATTAATCATTTGATCTGT 1713  
QY 1528 TTGAA-----AAAAGCTACATCATCGGAACACACAGTGTGGCCTTTGCAGATTTTAA 1581  
Db 1714 GATGAGAAACGAAAGAAATATGTATGTATGAGAGCTCATCTGGAATTTTGTGACTTCATG 1773  
QY 1582 ACTCCTCAGAAATGTGAGAAAGCCCATTTCTCAACCAAGAGGTGTTTCAAGAGACAGA 1641  
Db 1774 ACGAACCAGTCAACCACTGAGAGTAAACAGAAACAAGAGGGATCTTCACTGCACAGAGA 1833  
QY 1642 CAACCCAACTCGTGTCTCATGTACTGAGA 1671  
Db 1834 AACCACAGATGGCAGCTTCATTTTGGCA 1863

RESULT 12

US-10-421-175-3  
; Sequence 3, Application US/10421175  
; Publication No. US20030219414A1  
; GENERAL INFORMATION:  
; APPLICANT: Podsakoff, Gregory  
; APPLICANT: Watson, Gordon  
; APPLICANT: Couto, Linda B.  
; APPLICANT: Yang, Bin  
; TITLE OF INVENTION: RECOMBINANT ADENO-ASSOCIATED VIRUS VIRIONS FOR THE  
; TITLE OF INVENTION: TREATMENT OF LYSSOMAL DISORDERS  
; FILE REFERENCE: 0800-0021  
; CURRENT APPLICATION NUMBER: US/10/421,175  
; CURRENT FILING DATE: 2003-04-22  
; PRIOR APPLICATION NUMBER: US/09/715,858  
; PRIOR FILING DATE: 2000-11-14  
; NUMBER OF SEQ ID NOS: 16  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 3  
; LENGTH: 1947  
; TYPE: DNA  
; ORGANISM: Mus musculus  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(1947)  
US-10-421-175-3

Query Match

Best Local Similarity 9.4%; Score 159.6; DB 15; Length 1947;  
Matches 477; Conservative 0; Mismatches 444; Indels 21; Gaps 3;

QY 739 GAACCTGA AAAAGAGAGTACACTCTGACATCGGAATCAGAACGATCAGCTGGAGAGAG 798



Db 928 GAGTCTGTGACTGACTACTACACCCCTTCTATCGGGAATCGAACAAGTGGCTGTACAAAG 987  
QY 799 AAGAGGCTCTATCTGAAACGGGAAACCTGTCTTTTGAAGGGCTTTGGAAAGCAGAGGAA 858  
Db 988 AGCAAGTTCTCTATTAACGGGAAAGCCCTTCTATTTCCAAAGGGGTCAATTAAGCAGAGAT 1047  
QY 859 TTCCCGCTTCTGGGGCAGGACCTTTTATCCATGATGATAAAGAACTTCAACCTTCTG 918  
Db 1048 TCAGATATCCGAGGAAAGGCTTCGACTGCGCGCTGTGTAAGGATTTCAACCTGCTC 1107  
QY 919 AAGTGATCAACGCGAATTTCTTACAGACCTCTCATATCTTACAGTGAAGTGGCTG 978  
Db 1108 CGTTGGCTCGGGCAAAATTCCTTTCGTAACCAAGCCATATCCCTACTCAGAGGAGTACTT 1167  
QY 979 GATCTTGGCCGACAGACTCGGAATCCTTGTGATAGCAGAACCCCGCAGCTTG--TATC 1035  
Db 1168 CAGCTCTGTGACCGATACGGGATTTGTGTCATCGATGAGTGTCCCGGTGTGGCATTTGTG 1227  
QY 1036 ACAAGGTACCACTACATCCCGAGACTCAGAAAGTAGCAGAAACAACATAAGAGAATG 1095  
Db 1228 CTACTCAGAGTTTGGCAACGAGTCACTCCGCAACCACTAGAGTGTAGAGAGAGCTG 1287  
QY 1096 ATCGACAGACACAGAAACCATCCCACTGTGATCATGTGAGTGTGGCAACCAACAGAG 1155  
Db 1288 GTTCGCGCGGACAAATAATCAACCCGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1347  
QY 1156 TCCAACCATCCAGACGCGGAGGTTTCTCAAGCCCTTTATGAGACTGCCAATGAATG 1215  
Db 1348 TCTGCTCTGAAAACCCGCGCATATTAAGCGCTGATACCCACCAAAAGCCCTG 1407  
QY 1216 GATCGAACACGCGCGTGTGTGATGATGATGATGATGATGATGATGATGATGATGATG 1275  
Db 1408 GACCTCACCGTCCCGTGACCTTGTGAGCA-----ACGCCAATATGATGCA 1455  
QY 1276 GACGTGGCGTGAAGTACTTCAATCTGTGTGTGACAGGTAATAAGGCTGTGATC 1335  
Db 1456 GACCTGGGGCGCCCGTACGTGATGTATCTGTGTAACAGACTTCTTGTGTATCAT 1515  
QY 1336 TATCAGGGAAGATAGAAGAGACTTCAAGCTTGGAAAGACATAGAGAGCTTAT 1395  
Db 1516 GACTATGGGCAATTGGAGTGAATCAGCCACAGCTGAATGCGAGTTGGAAGTGTAT 1575  
QY 1396 GCAAGGCACAGAAAGCCCATTTTGTACAGAAATTCGTTGGGACGCGATAGCTGGCATC 1455  
Db 1576 AAGACGATCAGAAAGCCGATTAATCAGAGCAGATGATGAGCAGACGCAATCCAGGAGT 1635  
QY 1456 CACTACGATCACTCAATGTCTCCGAGAGTACCAAGCAGAGCTCGTTGAAAGACG 1515  
Db 1636 CACGAGACCGCTCGCATGTTCAGTAGAGTACCAAGAGCTGTTCGAGAGATTAAC 1695  
QY 1516 ATCAGGCTCTTTTGA-----AAAGACTACATCATCGGAACACACGTTGGCTTT 1569  
Db 1696 CATTGATGTCGATCAGAAAGTAAAGATACGTGTGAGAGCTCATCTGGAATTTT 1755  
QY 1570 GCAGATTTTAAGACTCTCTCAGATGTGAGAGACCATTTCAACCAAGAGGTGTTTC 1629  
Db 1756 GCCGACTTTCATGACGAACCGTCAACGCTGAGAGTAAATCGGAACAAGAGGATTTT 1815  
QY 1630 ACAAGAGACAGCAACCAACTCGTTGCTCATGTACTGAGA 1671  
Db 1816 ACTCGCAGAGACAGCCCAAACTTCGCTTTATTTGCGA 1857

## RESULT 13

US-10-120-145-7  
; Sequence 7, Application US/10120145  
; Publication No. US20030157684A1  
; GENERAL INFORMATION:  
; APPLICANT: Jefferson, Richard A.  
; APPLICANT: Kilian, Andrej  
; APPLICANT: Keese, Paul Konrad  
; TITLE OF INVENTION: MICROBIAL BETA-GLUCURONIDASE GENES, GENE PRODUCTS AND  
; TITLE OF INVENTION: USES THEREOF

FILE REFERENCE: 190106.405  
; CURRENT APPLICATION NUMBER: US/10/120,145  
; CURRENT FILING DATE: 2002-04-11  
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/149,727  
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-09-08  
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 60/058,263  
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-09-09  
; NUMBER OF SEQ ID NOS: 71  
; SOFTWARE: Patent In Ver. 2.0  
; SEQ ID NO 7  
; LENGTH: 1887  
; TYPE: DNA  
; ORGANISM: Bacillus sp.  
US-10-120-145-7

Query Match 8.4%; Score 141.4; DB 15; Length 1887;  
Best Local Similarity 45.4%; Pred. No. 3.9e-34;  
Matches 734; Conservative 0; Mismatches 822; Indels 60; Gaps 4;

QY 120 CCAGATCTGTGCTACGAAAGAGCCCTTCACTTACAAACCACTTCTTACGTTCCGAA 179  
Db 208 CAAGGAATCCCGCAACCATATCGGATATGTCTGTGTAAGAACGTTGACCGTCCGCGC 267  
QY 180 GNACTTTACAAACACATCAGACTTTACTTTGCTGCGGTGAACAGGACTCGGAGGT 239  
Db 268 CTATCTGAAGATGACCGTATCGTCTCCGCTTCCGCTCTGCACTCACMAAGCAATTGT 327  
QY 240 CTTCTCAACGAGAGAAAGTGGGAGAAATCAGATTTGATATCTTCCCTTCCAGATGA 299  
Db 328 CTATCTCAATGTGATGCTGTGTGAGACCAAGGCGGATTCTGCAATTGCAAGCGGA 387  
QY 300 TGTGACGGGGAAGTGAATCCCGAGAGAACCACTCAGGCTGTGTTGAGAACAGATT 359  
Db 388 AATCAACAACCTCGTGCCTGATGAGCATGAATCCGCTCACCGTCCGCTGACCAATCT 447  
QY 360 GAAAGTGGAGATTTCCCTCGAAGTTCAGACAGCGGCACTCACACCGTGGATTTT 419  
Db 448 CGACGATAGCAACCTCCCGTGGGCTGTACAGCGAGCGCCACGAGAGGCGCTCGGAAA 507  
QY 420 TGAAGTTTCCACTGCAACTTCCACTTCTTCCCTTCCGTTGAAATCATTAAGCCCTGT 479  
Db 508 AGTCAATTCGTAACAGCCGAACTTCCACTTCTTCACTATGACAGCGCTGACCGTCCGT 567  
QY 480 TGTGATAGATTCAACAGACCGCGAGATCTCGACATCTGGGTGACACAGATGATC 539  
Db 568 GAAATCTACAGACCCCGTTTACGTACGTGAGAGACATCTCGGTGTGACCGACTTCAA 627  
QY 540 TGAACCGGAGAAACTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 599  
Db 628 TGGCCCAACCGGACTGTGACCTATACGTGACTTCAAGGCAAAAGCCGAGACCGTGA 687  
QY 600 GGGACAGAGATGACGATCAAACTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 659  
Db 688 AGTGTGCTGCTGATGAG 747  
QY 660 ATTGTCGAGAGGAGTTCAATCTCGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 719  
Db 748 CGTGAAGATTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 807  
QY 720 TCTTATCTCTCAAGTGAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 779  
Db 808 AGTGAAGTGTGAAG 867  
QY 780 AACGATCAGCTGGAG 839  
Db 868 GACCGTGAAGTCAACGAGGCAAGTTCCTCATCAACAAACCGTTCTACTCAAGGG 927  
QY 840 CTTTGAAGACAGAGAGATTTCCCGTCTGGGAGAGAGAGAGAGAGAGAGAGAGAGAT 899  
Db 928 CTTTGAAGACATGAGAGACATCTCTATCAACGAGCGGCTTTAAGAGAGAGAGATGT 987  
QY 900 AAAAGACTTCAACCTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 959

Db 988 GATGGATTTCATATCTCAATGATCGGCGCCAAAGCTTCCGGAACCGACACTATCC 1047  
QY 960 TTACAGTGAAGAGTGGCTGGATCTTGCCGACAGACTCGGAATCTTTGTATAGAGAAC 1019  
Db 1048 GTACTCTGAAGAGTTGATGCGTCTTGGCGATCGGAGGGTCTGGTGTATCGACGAGAC 1107  
QY 1020 CCCGCACGTTGGTAT----- 1034  
Db 1108 TCCGGCAGTTGGCGTGCACTCAACTTCATGCGCACCAAGGACTCGGCGAAGGACGCA 1167  
QY 1035 ---CACAGGTACCACTACATCCGAGACTCAGAGATAGCAGAGACAACATAGAAG 1091  
Db 1168 GCGCGTCACTAGCTGGGAGAGATTGGACGTTGAGCACCATCAAGACGTTCTCCGTGA 1227  
QY 1092 AATGATCGACAGACAAGAACCATCCAGTGTGATCATGTGAGTGTGGCGAAGAAC 1151  
Db 1228 ACTGGTGTCTGTGACAAGAACCATCCAGCGTGTGATGTGAGCATCGCCAAAGAGGC 1287  
QY 1152 AGAGTCCAAACCATCCAGACGCGGAGGTTCTTCAAAGCCCTTTATGAGACTGCCAATGA 1211  
Db 1288 GCGCACTGAGGAGAGGCGCGTACGACTTCAAGCCGTTGTGAGACTGACCAAGGA 1347  
QY 1212 AATGGATCGA---ACACGCCCCGTTGTATGTGATGATGAGCATGATGACGACACAGAGAG 1268  
Db 1348 ACTGACCCACAGAACGCTCCGCTACGATCGTGTGTTGTGATGGCTACCCCGAGAC 1407  
QY 1269 AACAGAGACGTGGCGCTGAAGTACTTGACATCGTGTGTAACAGTACTACGCGCTG 1328  
Db 1408 GGACAAAGTCGCCGAAGTGA-----TTGACGTCAATCGCTCAATCGCTATAACGGATG 1461  
QY 1329 GTACATCTATCAGGGAAGATAGAGAG---GACTTCAAGCTCTGAAAAAGACATAGA 1385  
Db 1462 GTACTTGCATGCGGTGATCTGAAAGCGGCCAAAGTCCATCTCCGCGAGAAATTCACGC 1521  
QY 1386 AGAGCTCTATGCAAGGACAGAAAGCCCATCTTTGTACAGAAATTCGGTGGGACGCGAT 1445  
Db 1522 GTGGAACAAGCGTTGCCAGAAAGCCGATCATGATCACTGAGTACGGCGCAGACACCGT 1581  
QY 1446 AGCTGGCATCCACTACGATCCACTCAATGTTCTCCGAAGATCAAGCAGAGCTCGT 1505  
Db 1582 TGGCGGCTTTCACGACATTTGATCGATGATGTTCAACGAGAAATTCAGTCAAGTACTA 1641  
QY 1506 TGAAGAGAGATCAGGCTCTTTTGAAGAAAGATACATCATCGGAAACACAGCTGGGC 1565  
Db 1642 CCAGGCGAACCACGTCGTGTTGATGAGTTGAGAACTTCTGTGGTGAAGCAAGCGTGAA 1701  
QY 1566 CTTTGACAGATTTTAAGACTCTCTCAAGATGTGAGAAAGACCATTTTCAACCAAGGCTGT 1625  
Db 1702 CTTGCGGACTTCCGACCTCTCAAGGCGGTGATGCGGCTCCAAAGAAACAAGAGGCGGT 1761  
QY 1626 TTTTCAAGAGACAGACAAACCAACTCGTGTCTCATGTACTGAGAAAGCTGTGA 1681  
Db 1762 GTTCACTCGTGACCGCAAGCCGACGCTCGCGCGCAGCTCTTTCCGAGCGCTGA 1817

RESULT 14  
US-10-322-656-45

; Sequence 45, Application US/10322656  
; Publication No. US20030182691A1  
; GENERAL INFORMATION:  
; APPLICANT: Robert, Laurian  
; APPLICANT: Gledie, Stephen  
; APPLICANT: Foster, Elizabeth  
; TITLE OF INVENTION: Modification of Pollen Coat Protein Composition  
; FILE REFERENCE: 100210.01  
; CURRENT APPLICATION NUMBER: US/10/322,656  
; CURRENT FILING DATE: 2002-12-19  
; PRIOR APPLICATION NUMBER: 09/272,204  
; PRIOR FILING DATE: 1999-03-19  
; PRIOR APPLICATION NUMBER: 60/078,728  
; PRIOR FILING DATE: 1998-03-20  
; NUMBER OF SEQ ID NOS: 51  
; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 45  
; LENGTH: 3651  
; TYPE: DNA  
; ORGANISM: artificial  
; FEATURE:  
; OTHER INFORMATION: nucleotide sequence of the translational fusion in plasmid ATOG-4  
US-10-322-656-45

Query Match 8.4%; Score 141.4; DB 15; Length 3651;  
Best Local Similarity 45.4%; Pred. No. 5.7e-34;  
Matches 734; Conservative 0; Mismatches 822; Indels 60; Gaps 4;

QY 120 CCAGATCTGTGCTACGAGAGAACCCCTTACCTACAAACCCTTCTACGTTCCGA 179  
Db 1983 CAAGGAAATCCGCAACCATATCGGATATGTCTGGTACGAACGTGAGTTCAACGTCGCGC 2042  
QY 180 GNAACCTTCACAAAAACATCAGACTTACTTTGCTGCGGTGAACACGACTGCGAGT 239  
Db 2043 CTATCTGAAGGATCAGCGTATCGTCTCCGCTTCGCTCTGCAACTCACAAAGCAATTGT 2102  
QY 240 CTTCTCAACCGAGAGAAAGTGGAGAGAAATCATTGAATACCTTCCCTTCGAAGTGA 299  
Db 2103 CTATGTCAATGGTGAAGCTGTGCTGAGACCAAGGCGGATTCCTGCAATTCCGAAGCGGA 2162  
QY 300 TGTGACGGGAAAGTGAATCCGAGAGAAACGAATCAGGGTGTGTTGAGAACAGATT 359  
Db 2163 AATCAACAACCTCGCTGCGTATGATGATGAATCGCGTCAACGTCGCTGAGCAACATCCT 2222  
QY 360 GAAAGTGGAGAGATTCCCTCGAAGGTTCCAGACAGCGGACATCACACCGTGGATT 419  
Db 2223 CGACGATAGCACCTCTCCGCGTGGGCTGTACAGCGAGCGCCACGAAAGAGGCGCTCGGAAA 2282  
QY 420 TGAAGTTTCCACTGCAACTTTCGACTTCTTCCCTACCGTGGAAATCATAGGCGCTGT 479  
Db 2283 AGTCATTCTGAACAAGCCGAACCTTCACTTCAATATGCAAGGCTGCAACCGTCCGT 2342  
QY 480 TCTGATAGAGTTCAAGACACCGAGAGATCTGACATCTGGGTGACACGAGTGAATC 539  
Db 2343 GAAATCTACACGACCCCGTTTACGTACGTGAGAGACATCTCGGTGTGACCGACTTCAA 2402  
QY 540 TGAACCGAGAGAAACTTGAAGAAAGTGAAGTGAAGTGAAGTCTCAGAAAGCGGT 599  
Db 2403 TGGCCCAACCGGACTGTGACCTATACGTTGACCTTCAAGGCAAAAGCCGAGACCTGAA 2462  
QY 600 GGGACAGAGATGACGATCAAACTTGAGAGAGAGAAAAAGATTAGAACATCCAAAG 659  
Db 2463 AGTGTGCTGTGATGAGAGAGCAAAAGTGTGCAAGCAGAGGCGCTGAGCGTAA 2522  
QY 660 ATTGCTGAAGGGAATTCTATCTCGAAACGCGAGTCTGAGACCTCGAAGATCCATA 719  
Db 2523 CGTGAAGATTCGGAATGTCTCTGGAACCACTGAACACGTAATCTTACAGATCAA 2582  
QY 720 TCTTATCTCTCAAGGTGAAGTGAAGAAAGAGTACACTGTGACATCGAATCAG 779  
Db 2583 AGTGAAGTGTGAACGACGAGTGAACCATGATGTCTATGAAGAGCGTTCGCGTGC 2642  
QY 780 AACGATCAGTGGGACGAGAAAGGCTCTATCTGAACGGAACCTGTCTTTTGAAGG 839  
Db 2643 GACCGTGAAGTCAACGACGCGCAAGTCTCTCAATCAACAACAACCGTTCTACTTCAAGG 2702  
QY 840 CTTTGAAGCAGAGAAATCCCGTCTGCGGCGAGGCGACCTTTATCAATTGATGAT 899  
Db 2703 CTTTGGCAACATGAGACACACTCTATCAACGCGCGTGTAAAGAGAGCAATGT 2762  
QY 900 AAAAGACTTCAACTTCTGAAGTGAATCAACGGAATCTTTCAAGACCTTCATATCC 959  
Db 2763 GATGATTTCAATATCTCAATGATGATGCGGCAACAGCTTCGAGACCGCACTATCC 2822  
QY 960 TTACAGTGAAGTGGCTGATCTTGCGAGACTCGAATCTTGTGATGAGAGAGC 1019  
Db 2823 GTACTCTGAAGAGTTGATGCGTCTTGGAGATCGGAGGCTGTGATGATGAGAGAGC 2882

```
QY 1020 CCCGACGTTGGTAT----- 1034
Db 2883 TCCGGCAGTTGGCGTGCACTCACTTCATGCGCACCACGGACTCGGCGAAGCGCGA 2942
QY 1035 ---CACAAGGTACCATACAATCCCGAGACTCAGAAAGATAGCAGAGAACAACATAGAAG 1091
Db 2943 GCGCGTCAGTACCTGGGAGAAGATTGGAGCGTTTGAGCACCATCAAGACGTTCTCCGTGA 3002
QY 1092 AATGATCGACAGACACAAGAACCATCCCAAGTGTGATCATGTGAGTGTGGCGAAGAAC 1151
Db 3003 ACTGCTGTCTCTGACAGAAGAACCATCAAGCGTCTGATGTGAGCATCGCCAAAGAGGC 3062
QY 1152 AGAGTCCAAACCATCCAGACGCGGAGGTTTCTTCAAGCCCTTTATGAGACTGCCAATGA 1211
Db 3063 GCGGACTGAGAGAAGGCGCGGTACGAGTACTTCAAGCCGTTGGTGGAGCTGACCAAGGA 3122
QY 1212 AATGATCGA---ACACGCCCCCTGTCTCATGTGTGAGCATGATGAGCGCACCGAGAG 1268
Db 3123 ACTGACCCACAGAACGCTCCGCTACGATCGTGTCTTGTGATGGCTACCCCGAGAC 3182
QY 1269 AACAGAGACGTGCGCTGAGTACTTGCACATCGTCTGTGAAACAGGTACTACGGCTG 1328
Db 3183 GGAACAAGTCCCGCAACTGA-----TTGACGTCATCGCGCTCAATCGCTATTAACGATG 3236
QY 1329 GTACATCTATCAGGGAAGATAGAAGAG---GACTTCAAGCTCTGAAAAAGACATAGA 1385
Db 3237 GTACTTCGATGGCGGTGATCTCGAAGCGGCCAAAGTCCATCTCCCGCAGGAATTTACGC 3296
QY 1386 AGAGCTCTATGCAAGGCACAGAAAGCCCATCTTGTGTACAGAAATTCGGTGGAGCGCAT 1445
Db 3297 GTGGAACAAGCGTTGCCAGAAAGCCGATCATGATCACTGAGTACGGCGCAGACACCGT 3356
QY 1446 AGCTGGCATCCACTACATCCACTCAATGTCTCCGAAAGTACCAAGCAGAGCTCGT 1505
Db 3357 TGGGGGCTTTACAGACATGTGATCCAGTGTTCACCGAGAAATATCAAGTCAAGTACTA 3416
QY 1506 TGAAGAAGACGATCAGGCTCTTTGAAAAAGACTACATCATTCGGAACACACGTGTGGC 1565
Db 3417 CCAGGCCGAACCAAGTCTGTGATGAGTTGAGAACTTCGTGGGTGAGCAAGCGTGA 3476
QY 1566 CTTTGACAGATTTTAAGACTCTCAGAAATGTGAGAAGACCCCATTTCTCAACCAAGGCTGT 1625
Db 3477 CTTGCGGACTTCGCGACCTCTCAGGGCGGTGATGCGCTCCAAAGAAACAAGAGGCGT 3536
QY 1626 TTTCAACAAGACAGACAACCAAACTCGTCTCATGTACTGAGAAGACTGTGA 1681
Db 3537 GTTCACTCGTACCGCAAGCGGAAGCTCGCCGCGACGCTTTTCGCGAGCGCTGA 3592
```

## RESULT 15

US-10-322-656-48

```
/ Sequence 48, Application US/10322656
/ Publication No. US20030182691A1
/ GENERAL INFORMATION:
/ APPLICANT: Robert, Laurian
/ APPLICANT: Gledlie, Stephen
/ APPLICANT: Foster, Elizabeth
/ TITLE OF INVENTION: Modification of Pollen Coat Protein Composition
/ FILE REFERENCE: 100210.01
/ CURRENT APPLICATION NUMBER: US/10/322,656
/ PRIOR FILING DATE: 2002-12-19
/ PRIOR APPLICATION NUMBER: 09/272,204
/ PRIOR FILING DATE: 1999-03-19
/ PRIOR APPLICATION NUMBER: 60/078,728
/ NUMBER OF SEQ ID NOS: 51
/ SOFTWARE: PatentIn version 3.0
/ SEQ ID NO 48
/ LENGTH: 4084
/ TYPE: DNA
/ ORGANISM: artificial
/ FEATURE:
/ OTHER INFORMATION: nucleotide sequence of the translational fusion in plasmid EXLG-1
```

OTHER INFORMATION: (Example 25, Figure 30)  
US-10-322-656-48

Query Match 8.4%; Score 141.4; DB 15; Length 4084;  
Best Local Similarity 45.4%; Pred. No. 6.1e-34;  
Matches 734; Conservative 0; Mismatches 822; Indels 60; Gaps 4;

```
QY 120 CCAGATCTGTGTACGAGAAGGACCTTCACTACAAACCACCTTCTACGTTCCGAA 179
Db 2416 CAAGGAATCCGCAACCATATCGGATATGTTCTGTACGAACGTAAGTTCAACGTCGCGC 2475
QY 180 GNACTTTCAAAAAACATCAGACTTTACTTTGCTGCGGTGAACGCGACTGCGAGGT 239
Db 2476 CTATCTGAAGATCAGCGTATCGTCTCCGCTTCCGCTCTGCAACTCACAAAGCAATTGT 2535
QY 240 CTTCTCAACGAGAGAAAGTGGAGAGATCACATTTGATACCTTCCCTTGAAGTAGA 299
Db 2536 CTATGTCAATGTGTAGCTGTCTGTGAGCACAAGGCGGATTCTGCTCATTCGAACGGA 2595
QY 300 TGTACGGGGAAGTGAATCCGAGAGAACGAACCTCAGGCTGTTGTGAGACAGATT 359
Db 2596 AATCAACAACCTCGCTCGTGTATGGCATGAATCCGCTACCGCTCGGCTGACCAACATCT 2655
QY 360 GAAAGTGGAGAGATTCCCTCGAAGTTCCAGACAGCGGCACTCAACCGTGGATT 419
Db 2656 CGACGATAGCACCCCTCCGCTGGGTGTACAGCGAGCGCCACAGAGAGGCGCTCGGAAA 2715
QY 420 TGAAGTTTCCACCTGCAACCTTCACTTCTTCCCTACCGTGAATCATAGGCTGT 479
Db 2716 AGTCAATTCTGTAACAAGCCGAACCTTCACTTCTTCAACTATGACGCGCTGACACCTCGGT 2775
QY 480 TCTGATAGATTCAAGACACACCGGAGAGATCTGACATCTGGGTGACACGAGTAGTC 539
Db 2776 GAAATCTACACGACCCCGTTTACGTACGTGACATCTTCAAGGCAAAAGCCGAGACCTTCAA 2835
QY 540 TGAACCGGAGAGAACTTGAAGAGTGAAGTAGAAGTCTCAGAAAGCGGT 599
Db 2836 TGGCCCAACCGGAGCTGTGACCTTATACGTGACCTTCAAGGCAAAAGCCGAGACCTGAA 2895
QY 600 GGGACAGAGATGACGATCAAACTTGAAGAGAGAAAGAAAGATTGAACATCCAACAG 659
Db 2896 AGTGTGCTGTGATGAGAGAGCAAGTGTGCAAGCAGCAGAGGCGCTGAGCGTTAA 2955
QY 660 ATTCTGGAAGGAGTTCATCTTCGAAAACGCCAGGTTCTGAGCCTCGAAGATCCATA 719
Db 2956 CGTGAAGATTCCGAATGTCTCTCTGGAACCACTGMAACAGTATCTTACAGATCAA 3015
QY 720 TCTTTATCTCTCAAGGTGAACCTTGAAGAGAGAGTACACTCTGACATGGAATCAG 779
Db 3016 AGTGAAGTGTGAAGACAGCGGACTGACATGATGTCTATGAAGAGCCGTTGGCGGTGCG 3075
QY 780 AACGATCAGCTGGAGCAGAGAGGCTCTATCTGAACGGGAAACCTGTCTTTTGAAGG 839
Db 3076 GACCGTGGAAGTGAAGACAGCGGAAAGTTCTCTATCAACAAACACGTTCTACTTCAAGG 3135
QY 840 CTTTGAAGACAGAGGAATCCCGTCTGCGGCGAGGCGACCTTTATCAATTGATGAT 899
Db 3136 CTTTGGCAAACTGAGGACACTCTTATCAACGCGCGGTGCTTTAAGCAAGCAATGT 3195
QY 900 AAAAGACTTCAACCTCTGAAGTGATCAACGCGAATTTCTTCAAGACCTTCACTATCC 959
Db 3196 GATGATTTCAATATCTCAATGATCGCGCCCAACAGAGCTTCCGACCGCACATATCC 3255
QY 960 TTAAGTGAAGAGTGGCTGATCTTGCAGACAGACTCGGAATCCTGTGTAGACGAAGC 1019
Db 3256 GTACTCTGAAGAGTGTATGCGTCTTGCAGATCGGAGGCTGTGTGTGATCGACGAGAC 3315
QY 1020 CCCGACGTTGGTAT----- 1034
Db 3316 TCCGGCAGTTGGCGTGCACTCACTTCAATGCGCACCGGAGACTCGGCGAAGCGACGGA 3375
QY 1035 ---CACAAGGTACCATACAATCCCGAGACTCAGAAAGATAGCAGAGAACAATAAGAG 1091
```

```
Db      3376 GCGCGTCAGTACCTGGGAGAGAGATTCCGACGTTTGAGCACCATCAAGACGTTCTCCGTGA 3435
QY      1092 AATGATCGACAGACAGAACCATCCAGTGTGATCATGTGGAGTGGCGGAACCA 1151
      |||||
Db      3436 ACTGGTGTCTCGTGAAGAACCATCAAGCGTGTGATGTGAGCATCGCCAAAGAGGC 3495
QY      1152 AGAGTCCAACCATCCAGACGCGGAGGGTTTCTTCAAAGCCCTTATGAGACTGCCAATGA 1211
      |||||
Db      3496 GCGCACTGAGGAGAGGGCGCGTACAGTACTTCAAGCCGTTGGTGGAGCTGACCAAGGA 3555
QY      1212 AATGGATCGA--ACACGCCCCGTTGTCTATGTGTGAGCATGATGACGCCACAGAGAGAG 1268
      |||||
Db      3556 ACTCGACCCACAGAACCGTCCCGTCAAGATCGTGTGTTGTGATGGCTAACCCCGAGAC 3615
QY      1269 AACAAAGACGCTGGCGCTGAAGTACTTGCACATCGTCTGTGAAACAGTACTACGGCTG 1328
      |||||
Db      3616 GGACAAAGTCGCCGAAGCTGA-----TTGACGTCAATCGCGCTCAATCGCTATTAACGATG 3669
QY      1329 GTACATCTATCAGGGAAGGATAGAAGAG--GACTTCAAGCTCTGGAAGAAAGACATAGA 1385
      |||||
Db      3670 GTACTTCGATGGCGGTGATCTCGAAGCGGCCAAGTCCATCTCCGCCAGGAATTTCACGC 3729
QY      1386 AGAGCTCTATGCAAGGCCACAGAAAGCCCATCTTTGTCAAGAAATTCGGTCCGACCGAT 1445
      |||||
Db      3730 GTGGAACAAGCGTTGCCAGGAAAGCCGATCATGATCACTGAGTACGGCGCAGACACCGT 3789
QY      1446 AGCTGGCATCCACTACGATCCACCTCAAAATGTTCTCCGAAGAGTACCAAGCAGACTCGT 1505
      |||||
Db      3790 TCGGGGCTTTCACGACATTTGATCCAGTGTTCACCGAGGAATATCAAGTCGAGTACTA 3849
QY      1506 TGAAGAAGACGATCAGGCTCCTTTGAAAAAAGACTATCATTCGNAACACACGCTGGGC 1565
      |||||
Db      3850 CCAAGCGAACCACGTCGTGTTGATGAGTTGAGAACTTCGTGGTGAAGCAAGCGTGGAA 3909
QY      1566 CTTTGCAGATTTTAAGACTCCTCAGAAATGTGAGAGAACCCATTCTCAACCAAGGGTGT 1625
      |||||
Db      3910 CTTTCGCGGACTTCGCGAAGCTCTCAAGGCGGTGATGCGCTCCAAAGGAAACAAGAGGGCGT 3969
QY      1626 TTTCAACAAGAGACAGACAACCCAAAGCTGTTGCTCATGTACTGAGAAAGACTGTGA 1681
      |||||
Db      3970 GTTCACTCGTGACCGCAAGCCGAAGCTCGCCGCGCAGTCTTTTCGCGAGCGCTGA 4025
```

Search completed: January 24, 2005, 07:49:18  
Job time : 970 secs

This Page Blank (uspto)



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 24, 2005, 03:17:30 ; Search time 5564 Seconds  
(without alignments)  
11061.594 Million cell updates/sec

Title: US-09-936-759-14  
Perfect score: 1689  
Sequence: 1 atggttaagaccgcaacgaaa.....gaagactgtcgtgaggtc 1689

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 32822875 seqs, 18219865908 residues

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:  
1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_hic:\*  
4: gb\_est3:\*  
5: gb\_est4:\*  
6: gb\_est5:\*  
7: gb\_est6:\*  
8: gb\_gss1:\*  
9: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	174	10.3	1377	3	CR593823 full-length
2	174	10.3	1853	6	CD014093 90135266
3	174	10.3	1995	6	CD014092 90134967
4	161.6	9.6	1124	6	CD503076 CDA60-C07
5	161.2	9.5	1138	6	CD503098 CDA60-D07
6	159.6	9.4	2274	3	AK041058 Mus muscu
7	150.6	8.9	816	7	CK467358 938649 MA
8	144.4	8.5	906	5	BQ941196 AGENCOURT
9	142.6	8.4	878	7	CN155220 942501 MA
10	141.6	8.4	1051	5	BX363460 BX363460
11	132.6	7.9	767	6	CF255373 mdm127 c
12	131	7.8	846	6	CA453907 AGENCOURT
13	129.8	7.7	900	6	CA489317 AGENCOURT
14	129.8	7.7	914	6	CB203472 AGENCOURT
15	128.2	7.6	862	7	CO394571 AGENCOURT
16	126.4	7.5	2473	3	AY321342 Rattus no
17	121	7.2	857	5	BX745933 BX745933
18	117.4	7.0	1287	6	CD503099 CDA60-D07
19	117.2	6.9	802	5	BU357212 603474094
20	116	6.9	725	5	BU315788 603851444
21	114.2	6.8	724	5	BM441165 BM441165
22	112.6	6.7	771	4	BG121498 602352830
23	111.6	6.6	790	7	CO559438 AGENCOURT
24	110.8	6.6	740	7	CF521612 AGENCOURT

C	25	109.6	6.5	752	4	BU733141	BU733141
	26	109.4	6.5	1268	6	CD503077	CD503077 CDA60-C07
	27	107.6	6.4	673	6	CB841568	CB841568 M15E-2121
	28	107.4	6.4	791	5	BU239978	BU239978 603323761
	29	107.2	6.3	915	5	BQ678153	BQ678153 AGENCOURT
	30	106.6	6.3	543	5	BQ360163	BQ360163 QV0-OT003
	31	106.6	6.3	915	4	BG568800	BG568800 602588157
	32	106.4	6.3	975	4	BG285429	BG285429 602406582
	33	105.6	6.3	616	2	BE179118	BE179118 RC0-HT061
	34	105.4	6.2	504	5	BU917201	BU917201 EST143 Bo
	35	105.2	6.2	736	5	BP454056	BP454056 BP454056
	36	104.6	6.2	757	4	BI851705	BI851705 603377940
C	37	104	6.2	877	7	CF406520	CF406520 CH3#042 G
	38	101.8	6.0	1023	5	BX363459	BX363459 BX363459
	39	101.6	6.0	773	4	CF748482	CF748482 UI-M-HJ0-
	40	101	6.0	777	4	BG741172	BG741172 602631852
	41	100.8	6.0	667	7	CN264045	CN264045 170004243
	42	100.6	6.0	548	2	BE751040	BE751040 202745 MA
	43	100.2	5.9	799	5	BU366987	BU366987 603585124
	44	100	5.9	701	5	BX611242	BX611242 BX611242
C	45	99.6	5.9	921	9	CL486845	CL486845 SATL_443_

ALIGNMENTS

RESULT 1  
LOCUS CR593823 1377 bp mRNA linear HTC 21-JUL-2004  
DEFINITION full-length cDNA clone CSODL001YM21 of B cells (Ramos cell line)  
Cot 25-normalized of Homo sapiens (human).  
ACCESSION CR593823  
VERSION CR593823.1 GI:50474630  
KEYWORDS HTC; CNSLT\_cDNA.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE 1 (bases 1 to 1377)  
AUTHORS Li,W.B., Gruber,C., Jeesee,J. and Polayes,D.  
TITLE Full-length cDNA libraries and normalization  
JOURNAL Unpublished  
REMARK Contact : Feng Liang Email : fliang@lifetech.com URL :  
http://fulllength.invitrogen.com/Invitrogen Corporation 1600  
Faraday Avenue  
2 (bases 1 to 1377)  
REFERENCE Genoscope.  
AUTHORS Direct Submission  
TITLE Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :  
JOURNAL BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)  
COMMENT - Web : www.genoscope.cns.fr)  
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime  
end enriched, double-strand cDNA was digested with Not I and cloned  
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library  
was normalized. Library was constructed by Life Technologies, a  
division of Invitrogen.  
FEATURES  
Location/Qualifiers  
1..1377  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CSODL001YM21"  
/tissue\_type="B cells (Ramos cell line) Cot 25-normalized"  
/plasmid="pCMVSPORT\_6"  
ORIGIN  
Query Match 10.3%; Score 174; DB 3; Length 1377;  
Best Local Similarity 52.2%; Pred. No. 7.3e-40;  
Matches 493; Conservative 0; Mismatches 430; Indels 21; Gaps 4;  
QY 737 TGGAACTTGAAGAAAAGACGAGTACACTCTGGACATCGGAATCAGAAGATCAGCTGGAGCG 796  
DB 215 TGGGGCCTGTGTCTGACTTCTTACACACTCCCTGTGGGATCCGACACTGTGGCTGTACCA 274

```

OY 797 AGAAGAGCTCTATCTGAACGGGAACCTGTCTTTTGAAGGCTTTGGAAGACGAGG 856
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 275 AGAGCCAGTCTCTATCAATGGGAACCTTTCTATTTCCACGGTGTCAACAAGCATGAGG 334
OY 857 AATTCCCCGTTCTGGGGCAGGGACCTTTTATCCATGTATGATAAAGACTTCAACCTTC 916
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 335 ATGGGACATCCGAGGGAAGGCTTGAAGTGGCGCTGTGTGAAGAGACTTCAACCTGC 394
OY 917 TGAAGTGAATCAAGCGAATCTTTCAGAGACTCTCACTATCTTACAGTGAAGAGTGGC 976
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 395 TTGCTGGCTTGTGTGCAACGCTTTCGTACCAAGCACTACCCCTATGACAGGAAGTGA 454
OY 977 TGGATCTTGGCAGAGACTCGGAATCCTTGTGATAGAGGAAGCCCGCAGCTTGTATCA 1036
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 455 TGCAGATGTGTGACCGCTATGGGATGTGTGATGATGATGTGTCCGCGGCTGG 514
OY 1037 CAAGGTACCA---CTACAATCCGAGACTCAGAGATAGCAGAAAGACAATAAGAGAA 1093
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 515 CGCTGCCGAGTCTTCAACAACGTTTCTGTGATCACCACATGCAGGTGATGGAAGAG 574
OY 1094 TGATCGACAGACACAAGAACCAATCCAGTGTGATCATGTGAGTGTGCGGAACGACAG 1153
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 575 TGGTGCCTAGGGAACAAGAACCCCGCGTGTGATGTGTGTGTGTGCGCAACGAGCTG 634
OY 1154 AGTCCAACCATCCAGACGCGGAGGTTTCTTCAAGCCCTTATGAGACTGCCAATGAAA 1213
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 635 CGTCCACCTAGAAATCTGTGCTGTAATCACTGAAGATGTGATTCGCTCACCAATCCT 694
OY 1214 TGGATCGAACACGCCCCGTTGTTCATGTGAGCATGTGACGCAACGAGAGAGAACAA 1273
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 695 TGGACCCCTCCGCGCTGTGACCTTGTGAGCACTTAATATGACAG----- 742
OY 1274 GAGACGTGGCGCTGAAGTACTTGCACATCGTGTGTGAACAGTACTACGGCTGTACA 1333
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 743 CAGACAAAGGGGCTCCGTATGTGATGTGATCTGTTGAACAGTACTACTCTTGTATC 802
OY 1334 TCTATCAGGGAAGATAGAAAGAGACTTCAAGCTCTGGAAGAAAGACATAGAAGCTCT 1393
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 803 ACGACTACGGGACCTGAGTTGATTCACTGACAGCTGGCCACCCAGTTGAGAAGTGGT 862
OY 1394 ATGCAAGGCACAGAAAGCCCACTTTTGTCAAGAAATTCGGTGGGACGCGATAGCTGCA 1453
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 863 ATAAAGATATCAGAAAGCCCACTTATTCAGAGCGAGTATGAGCAGAAAGATTCAGAGGT 922
OY 1454 TCCACTACGATCCACCTCAAAATGTTCTCCGAAGAGTACCAAGCAGAGCTGTTGAAAAG- 1512
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 923 TTCACCAAGATCCACCTCTGATGTTCATCTGAAGAGTACCAAGAAAGTCTAGAGCAGT 982
OY 1513 ACGATCAGGCTCTTTTGAAGAA---AGACTACATCATCGGAACACAGCTGTGGGCT 1567
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 983 ACCATCTGGGTCTGATCAAAAACGAGAAATACGTGTTGAGAGCTCATTTGAATTT 1042
OY 1568 TTGAGATTTTAAGACTCTCAGAAATGTGAGAAAGACCAATCTCAACCACAAGGCTGTTT 1627
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1043 TTGCCGATTTTCACTGACTGAACAGTCAACCAAGAGAGTGTGGGAATTAAGGGAATCT 1102
OY 1628 TCACAAGAGACAGAACCAAACTCGTTGCTCATGTAAGTGA 1671
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1103 TCACCTCGGACAGAACCAAAAGTGCAGCGTTCCTTTGCGA 1146

```

## RESULT 2

```

LOCUS      CD014093      1853 bp      mRNA      linear      EST 21-OCT-2003
DEFINITION 90135266 single gene library Homo sapiens cdna, mRNA sequence.
ACCESSION  CD014093
VERSION     CD014093.1  GI:37777622
KEYWORDS   EST.
SOURCE      Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```

```

REFERENCE  1 (bases 1 to 1853)
AUTHORS   Jin, P., Fu, G.K., Wilson, A.D., Yang, J., Chien, D., Hawkins, P.R.,
            Au-Young, J., and Stuve, L.L.
TITLE      PCR isolation and cloning of novel splice variant mRNAs from known
            drug target genes
JOURNAL    Genomics 83 (4), 566-571 (2004)
COMMENT    Contact: Jin, P.
            Incyte Corporation
            3160 Porter Drive, Palo Alto, CA 94304, USA
            Tel: 650 621 8639
            Fax: 650 621 8965
            Email: pj@incyte.com.
FEATURES   location/Qualifiers
            source
            1..1853
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone_lib="Single gene library"
            /note="Vector: pDrive Cloning Vector; RT-PCR was performed
            using gene-specific primers flanking the open-reading
            frame. PCR products were subcloned into pDrive cloning
            vector and sequenced completely using M13 forward and
            reverse primers. Sequencing gaps were closed by
            re-sequencing using primers flanking the gapped areas."
ORIGIN

```

```

Query Match      10.3%; Score 174; DB 6; Length 1853;
Best Local Similarity 52.2%; Pred. No. 8.2e-40;
Matches 493; Conservative 0; Mismatches 430; Indels 21; Gaps 4;

OY 737 TGAACCTGAAAAAGACGAGTACACTCTGACATCGGAATCAGAACGATCAGCTGGAGC 796
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 624 TGGGCGCTGTGTGACTTCTACACACTCCCTGTGGGATCCGACATGTGCTGTACCA 683
OY 797 AGAAGAGCTCTATCTGAACGGGAACCTGTCTTTTGAAGGCTTTGGAAGCAGAGG 856
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 684 AGAGCAGTTCTCATCAATGGGAACCTTTCTATTTCCACGGTGTCAACAAGCATGAGG 743
OY 857 AATTCCCCGTTCTGGGGCAGGGACCTTTATCCATTTGATGATAAAGACTTCAACCTTC 916
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 744 ATGGGACATCCGAGGGAAGGGCTTGCAGTGGCGCTGTGTGAAGAGACTTCAACCTGC 803
OY 917 TGAAGTGAATCAAGCGAATTTCTTCAAGACCTCTCACTATCCTTACAGTGAAGAGTGC 976
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 804 TTGCGTGGCTTGTGTGCCAACGCTTTCGTACACAGCCACTACCCCTATGACAGGAAGTGA 863
OY 977 TGGATCTTGGCAGACACTCGGAATCCTTGTGATAGAGGAAGCCCGCAGCTGTGATCA 1036
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 864 TGCAGATGTGTGACCGCTATGGGATTTGTGTCATCACCACATGCAGGTGATGGAAGAG 923
OY 1037 CAAGGTACCA---CTACAATCCGAGACTCAGAAAGATAGCAGAAAGACAATAAGAGAA 1093
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 924 CGTCCGAGTCTTCAACAACGTTTCTGTGATCACCACATGCAGGTGATGGAAGAG 983
OY 1094 TGATCGACAGACACAAGAACCAATCCAGTGTGATCATGTGAGTGTGGGAAAGAACAG 1153
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 984 TGGTGCCTAGGGAACAAGAACCAACCCGCGTGTGATGTGTGTGTGCGCAAGAGCTG 1043
OY 1154 AGTCCAACCATCCAGACGCGGAGGTTTCTTCAAGCCCTTATAGAGACTGCCAATGAAA 1213
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1044 CGTCCACCTAGAAATCTGTGGCTACTACTGAAGATGTGATCGCTCACACCAATCCT 1103
OY 1214 TGGATCGAACACGCCCCGTTGTGATGTGAGCATGATGAGCGCACAGACGAGAGAACAA 1273
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1104 TGGACCCCTCCGCGCTGTGACCTTGTGAGCAACTTAATATGACAG----- 1151
OY 1274 GAGACGTGGCGCTGAAGTACTTGCACATCGTGTGTGAAACAGTACTACGGCTGTACA 1333
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1152 CAGACAAAGGGGCTCCGTATGTGATGTGATCTGTTGAACAGTACTACTCTTGTATC 1211
OY 1334 TCTATCAGGGAAGATAGAGAAGACTTCAAGCTCTGGAAGAAAGACATAGAAGCTCT 1393
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1212 ACGACTACGGGCACCTGAGTTGATTCAGCTGCAGCTGGCCACCCAGTTTGAAGAAGTGGT 1271

```

QY 1394 ATGCAAGGCACAGAAAGCCCATCTTGTGCACAGAAATTCGGTGCAGACGCGATAGCTGGCA 1453  
DB 1272 ATAGAAGTATCAGAAAGCCCATTTTCAGAGCGAGTATGGAGCAGAAACGATTGCAGGGT 1331  
QY 1454 TCCAGTACGATCCACCTCAAAATGTTCTCCGAAGAGTACCAAGCAGAGCTCGTTGAAAAG- 1512  
DB 1332 TTCACCAAGATCCACTCTGATGTTCACTGAAGAGTACCAGAAAAGTCTGCTAGACAGT 1391  
QY 1513 ACGATCAGGCTCCTTTTGAAAAA-----AGACTACATCATCCGAAACACAGCTGTGGCCCT 1567  
DB 1392 ACCATCTGGGTCTGGATCAAAAACGAGAAAATACGTGTTGGAGAGCTCATTTGGAATT 1451  
QY 1568 TTGCAGATTTTAAGACTCCTCAGAGATGTGAGAGACCCATTCTCAACCAAGGCTGTT 1627  
DB 1452 TTGCCGATTTTCATGACTGAACAGTCAACCGACGAGAGTGTGGGGAATAAAAAGGGATCT 1511  
QY 1628 TCACAAGAGACAGCAACCCAACTCGTTGCTCATGTACTGAGA 1671  
DB 1512 TCACTCGGACAGACAAACCAAAAGTGCAGCGTTCCTTTTGGCA 1555

RESULT 3  
CD014092 1995 bp mRNA linear EST 21-OCT-2003  
LOCUS 90134967 Single gene library Homo sapiens cDNA, mRNA sequence.  
DEFINITION CD014092  
ACCESSION CD014092  
VERSION CD014092.1 GI:37777621  
KEYWORDS EST.

SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 1995)

AUTHORS Jin, P., Fu, G.K., Wilson, A.D., Yang, J., Chien, D., Hawkins, P.R.,  
Au-Young, J. and Stuve, L.L.

TITLE PCR isolation and cloning of novel splice variant mRNAs from known  
drug target genes

JOURNAL Genomics 83 (4), 566-571 (2004)

COMMENT Contact: Jin, P.  
Incyte Corporation  
3160 Porter Drive, Palo Alto, CA 94304, USA  
Tel: 650 621 8639  
Fax: 650 621 8965  
Email: pjlin@incyte.com.

FEATURES  
source 1. 1995  
Location/Qualifiers  
1. 1995  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone\_lib="Single gene library"  
/note="Vector: pDrive Cloning Vector; RT-PCR was performed  
using gene-specific primers flanking the open-reading  
frame. PCR products were subcloned into pDrive Cloning  
Vector and sequenced completely using M13 forward and  
reverse primers. Sequencing gaps were closed by  
re-sequencing using primers flanking the gapped areas."

ORIGIN

Query Match 10.3%; Score 174; DB 6; Length 1995;  
Best Local Similarity 52.2%; Pred. No. 8.4e-40;  
Matches 493; Conservative 0; Mismatches 430; Indels 21; Gaps 4;

QY 737 TGGAACTTGAAAAAGCAGTACTGTGACATCGGAATCAGACGATCAGCTGGAGC 796  
DB 766 TGGGGCTGTGTGACTTCTACACTCCCTGTGGGATCCGCACTGTGCTGCACCA 825  
QY 797 AGAAGAGGCTTATCTGAACGGGAACTGTCTTTTGAAGGCTTTGAAAGCAGAG 856  
DB 826 AGAGCCAGTTCTCATCAATGGGAACTTTCTATTTCACGCTGTCAACAAGCATGAGG 885  
QY 857 AATTCCCGTTTGGGGCAGGCGCACTTTTATCCATGTGATGATAAAGACTTCACTTC 916

DB 886 ATCGGACATCCGAGGGAGGGCTTCGACTGGCCGCTGCTGTGAAGACTTCAACTTC 945  
QY 917 TGAAGTGATCAACGGGAATTTCTTACAGACCTCTCACTATCTTACAGTGAAGTGGC 976  
DB 946 TTCGCTGGCTTGGTGCACACGCTTTTCCGTACCAAGCCACTACCTTATGACAGAGTGA 1005  
QY 977 TGGATCTTGGCCGACAGACTCGGAATCCTTGTGATAGAGAAAGCCCGCAGCTTGATCA 1036  
DB 1006 TGCAGATGTGTGACCGCTATGGGATGTGTGTATCGATGAGTGTCCCGCGTGGCCCTGG 1065  
QY 1037 CAAGGTACCA---CTACATCCGAGACTCAGAAATAGCAGAAACAAATAGAAGAA 1093  
DB 1066 CGCTGCCGAGTTCTTCAACAAGTTTCTCTGCATCACACATGCAGGTGATGAAGAAG 1125  
QY 1094 TGATGCACAGACAAAGAACCATTCCAGTGTGATCATGTGAGTGTGGCGAAACGAC 1153  
DB 1126 TGGTGCGTAGGGACAAGAACCAACCCCGGCTGTGATGTGTGTGTGGCCAAAGCCTG 1185  
QY 1154 AGTCCAACCATCCAGACGCGGAGGTTTCTTCAAAAGCCCTTTATGAGACTGCCAATGAAA 1213  
DB 1186 CGTCCACCTAGAAATCTGTGCTACTACTTGAAGATGTGTATCGCTCACACCAAAATCCT 1245  
QY 1214 TGGATCGAACAGCCCCGTTGTGATGTGAGCATGTATGAGCGCACAGACGAGAAACAA 1273  
DB 1246 TGAACCCCTCCCGGCTGTGACCTTTGTGAGCAACTTAATCTATGACAG----- 1293  
QY 1274 GAGAGGTGGCGTGAAGTACTTTCAGATCGTGTGTGAAACAGTACTACGGCTGTACA 1333  
DB 1294 CAGACAAAGGGGCTCCGTATGTGATGTGATGTGATGTGTTGAACAGTACTACTCTGTGATC 1353  
QY 1334 TCTATCAGGGAAGATAGAAGAAGACTTCAAGCTCTGAAAAAAGACATAGAAGACTCT 1393  
DB 1354 ACGACTACGGGACCTGAGTGTATTCAGCTGCAGCTGCGCACCAAGTTTGAAGACTGGT 1413  
QY 1394 ATGCAAGGCACAGAAAGCCCATCTTGTGTCAAGAAATTCGGTGGGACGCGATAGCTGGCA 1453  
DB 1414 ATAGAAGTATCAGAAAGCCCATTTATTCAGAGGAGTATGAGCAGAAACGATGTCAAGGT 1473  
QY 1454 TCCACTAGATCCACCTCAATGTTCTCCGAGAGTACCAGCAGAGCTCGTTGAAAAG- 1512  
DB 1474 TTCACCAAGATCCACTCTGATGTTCACTGAAGAGTACCAGAAAAGTCTGCTAGACAGT 1533  
QY 1513 ACGATCAGGCTCCTTTGAAAAA-----AGACTACATCATCGGAACACAGTGTGGCCCT 1567  
DB 1534 ACCATCTGGGTCTGGATCAAAAACGAGAAAATACGTGTTGAGAGACTCATTTGGAATT 1593  
QY 1568 TTGCAGATTTTAAGACTCCTCAGAAATGTGAGAGACCCATTCTCAACCAAGGCTGTT 1627  
DB 1594 TTGCCGATTTTCATGACTGAACAGTCAACGACGAGAGAGTGTGGGAAATAAAAAGGGATCT 1653  
QY 1628 TCACAAGAGACAGCAACCCAACTCGTTGCTCATGTACTGAGA 1671  
DB 1654 TCACTCGGACAGACAAACCAAAAAGTGACAGCGTTCCTTTTGGCA 1697

RESULT 4  
CD503076 1124 bp mRNA linear EST 12-JUN-2003  
LOCUS CDA60-C07.x1d-t SHGC-CDA Gasterosteus aculeatus cDNA clone  
DEFINITION CDA60-C07 5', mRNA sequence.  
ACCESSION CD503076  
VERSION CD503076.1 GI:31432977  
KEYWORDS EST.  
SOURCE Gasterosteus aculeatus (three spined stickleback)  
ORGANISM Gasterosteus aculeatus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
Acanthomorpha; Acanthopterygii; Percomorpha; Gasterosteiformes;  
Gasterosteidae; Gasterosteus.  
REFERENCE 1 (bases 1 to 1124)  
AUTHORS Kingsley, D.M., Peichel, C., Balabandha, S., Grimwood, J., Dickson, M.,  
Schmutz, J. and Myers, R.M.  
TITLE Expressed sequence tags from Gasterosteus aculeatus

JOURNAL  
COMMENT

Unpublished (2003)  
Contact: Kingsley, DM  
HHMI and Department of Developmental Biology  
Stanford University School of Medicine  
Beckman Center B300, 279 Campus Drive, Stanford, CA 94305-5329, USA  
Tel: 650 725 5954  
Fax: 650 725 7739  
Email: kingsley@cmgm.stanford.edu  
Plate: 60  
High quality sequence stop: 782.

## FEATURES

## source

1. .1124  
Location/Qualifiers  
/organism="Gasterosteus aculeatus"  
/mol\_type="mRNA"  
/strain="Salinas river, CA"  
/db\_xref="taxon:69293"  
/clone="CDA60-C07"  
/sex="mixed male and female"  
/tissue\_type="heads and internal organs combined"  
/dev\_stage="adult"  
/clone\_lib="SHGC-CDA"  
/note="Vector: lambda ZAP Express/PBK-CMV; Site 1: EcoRI (5' adaptor); Site 2: XhoI (3' linker primer); The mixed organ cDNA library was generated using the ZAP-cDNA method by Stratagene. First strand cDNA synthesis was primed with a 50 bp linker primer containing an oligo dT sequence preceded by a synthetic XhoI site. 5 prime adaptors were used containing an EcoRI cohesive end. The finished cDNAs were inserted in to the ZAP express vector undirectionally in the sense orientation with respect to the lacZ promoter of PBK-CMV. An amplified library was prepared from approximately 3 million primary clones in the lambda ZAP Express vector. In vivo excision was then used to generate individual PBK-CMV phagemid clones for EST sequencing."

## ORIGIN

Query Match 9.6%; Score 161.6; DB 6; Length 1124;  
Best Local Similarity 51.5%; Pred. No. 3.4e-36;  
Matches 461; Conservative 0; Mismatches 414; Indels 21; Gaps 3;

QY 748 AAGACGAGTACACTCTGGACATCGGAATCAGAACGATCGTGGACGAGAAGAGCTC 807  
DB 71 AAGCAGGTGACACTCTACAGTGGCATCGGACGCTGACGTTACAGCAACCACTTC 130  
QY 808 TATCTGAACGGGAAACCTGTCTTTTGAAGGCTTTGAAAGCAGAGAAATCCCGTT 867  
DB 131 CTCATCAACAAAGCCCTTCTACTTCCACGAGTAATATAACACGAGACTCTGATATT 190  
QY 868 CTGGGGCAGGGCACTTTATTCATGTATGATATAAAGACTTCAACCTTCTGAAGTGATC 927  
DB 191 CGAGGCAAGGCCCTGACTGGCCCTCATGTGTGAAGACTTTAATTATGAAGTGTTG 250  
QY 928 AACGGGAATCTTTCAGGACCTCTCACTATCTTACAGTGAAGAGTGCTGATCTTGC 987  
DB 251 GGGGCCAATCTGTTCCGACCAAGCCACTACCTTATGACAGAGATCTGCAGATGTGT 310  
QY 988 GACAGACTGGGAATCTTGTGATAGACGAAGCCCGCAGCTGTGATACA--AGGTAC 1044  
DB 311 GACCGCCATGGCATCGTGTGATAGACGAGTGCCCGGCGTGGCATAAAGACATTCGC 370  
QY 1045 CACTACAATCCCGAGACTAGAGAAGATAGACAAGACAATTAAGAGATGATCGACAGA 1104  
DB 371 AGTTTGAAGACGCTCTTAACCATCACTGTGTCATGACGAGCTGTACTCGG 430  
QY 1105 CACAAGAACCATCCAGTGTGATCATGTGAGTGTGGGAAAGAACGAGATCCAAACAT 1164  
DB 431 GACAAGAACCATCCCTGTGTGTGATGTGTCAGTGGCCAATGAGCCGCTGCAGAGATG 490  
QY 1165 CCAGACGGGAGGGTTTCTTCAAGCCCTTTATGAGACTGCCAATGAATGATCGACACA 1224  
DB 491 CCTCCTGTGATTACTATTCAAAACCTTGATAAACATACCAAGAAATTGATCCAAAC 550

QY 1225 CGCCCCGTTGTCATGTGAGCATGATGAGACCGCACAGAGAGAAGACAGAGCTGGCG 1284  
DB 551 CGGCCCGT-----CACTTTATCACAGACAGTAATATGATCCAGGATAAAGG 598  
QY 1285 CTGAAGTACTTCGACATCGTCTGTGTGAACAGTACTACGGCTGTGTATCATCTACAGGA 1344  
DB 599 GCTCCCTACGTGACGTCATCTGCGTAAACAGTACTCTCCTGGTACCATGACCCCGGC 658  
QY 1345 AGATAGAAAGAGACTTCAAGCTTGAAAAAGACATAGAAGAGCTTATGCAAGGAC 1404  
DB 659 CACCCGAGGTATCCCATCCAGCTCAACACTCATGTTGAGAACTGTACGAAAGTAC 718  
QY 1405 AGAAGCCATCTTTGTCAAGAAATCGGTGCGAGCGGATAGCTGTCATCCATCGAT 1464  
DB 719 CAGAAACCATCATCCAGAGCGAATACGAGCGGATGCGGCTTCCACAGTAT 778  
QY 1465 CCACCTCAATGTCTTCCGAGAGTACCA-----AGCAGAGCTGTTGAAAAGCATC 1518  
DB 779 CCACCGTGATGTTACTGAGAGTACCAAGATTAGTCTGCAGAGCTACCAACAGTG 838  
QY 1519 AGCTCCTTTTGAAGAAAGACTACATCATCGAACAACAGTGTGGCCTTGACATTTT 1578  
DB 839 TTGCACAGAAAGAGACAGTACATCATCGGCAACTCATCTGAACTTTCAGACTTC 898  
QY 1579 AAGACTCTCAGAAATGTGAGAAAGCCATTTCAACCAAGGGTGTTCACAG 1634  
DB 899 ATGACCCACACAGGATCATGCGGTGTGTGGGAACAAAAAGGTGTCTCAGCAG 954

RESULT 5  
CD503098 1138 bp mRNA linear EST 12-JUN-2003  
LOCUS CDA60-D07.x1d-t SHGC-CDA Gasterosteus aculeatus cDNA clone  
DEFINITION CDA60-D07 5', mRNA sequence.  
ACCESSION CD503098  
VERSION CD503098.1 GI:31433163  
KEYWORDS EST.  
SOURCE Gasterosteus aculeatus (three spined stickleback)  
ORGANISM Gasterosteus aculeatus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
Acanthomorpha; Acanthopterygii; Percomorpha; Gasterosteiformes;  
Gasterosteidae; Gasterosteus.  
1 (bases 1 to 1138)

REFERENCE  
AUTHORS Kingsley, D.M., Peichel, C., Balabhadra, S., Grimwood, J., Dickson, M.,  
Schmutz, J. and Myers, R.M.  
TITLE Expressed sequence tags from Gasterosteus aculeatus  
JOURNAL Unpublished (2003)  
COMMENT Contact: Kingsley, DM  
HHMI and Department of Developmental Biology  
Stanford University School of Medicine  
Beckman Center B300, 279 Campus Drive, Stanford, CA 94305-5329, USA  
Tel: 650 725 5954  
Fax: 650 725 7739  
Email: kingsley@cmgm.stanford.edu  
Plate: 60  
High quality sequence stop: 782.

FEATURES  
source  
1. .1138  
Location/Qualifiers  
/organism="Gasterosteus aculeatus"  
/mol\_type="mRNA"  
/strain="Salinas river, CA"  
/db\_xref="taxon:69293"  
/clone="CDA60-D07"  
/sex="mixed male and female"  
/tissue\_type="heads and internal organs combined"  
/dev\_stage="adult"  
/clone\_lib="SHGC-CDA"  
/note="Vector: lambda ZAP Express/PBK-CMV; Site 1: EcoRI (5' adaptor); Site 2: XhoI (3' linker primer); The mixed organ cDNA library was generated using the ZAP-cDNA method by Stratagene. First strand cDNA synthesis was primed with a 50 bp linker primer containing an oligo dT sequence

preceded by a synthetic XhoI site. 5 prime adaptors were used containing an EcoRI cohesive end. The finished cDNAs were inserted in to the ZAP express vector unidirectionally in the sense orientation with respect to the lacZ promoter of PBK-CMV. An amplified library was prepared from approximately 3 million primary clones in the lambda ZAP Express vector. In vivo excision was then used to generate individual PBK-CMV phagemid clones for EST sequencing."

ORIGIN

Query Match	9.5%;	Score 161.2;	DB 6;	Length 1138;
Best Local Similarity	51.9%;	Pred. No. 4.5e-36;		
Matches	473;	Conservative 0;	Mismatches 418;	Indels 21; Gaps 4;
QY	748	AAAGACGAGTACTCTTGACATCGAATCAGACGCTGGGACGAGAAGGCTC	807	
DB	71	AACGACGTGTACTCTTACCAGTCGGCATCCGACGGTCGACGTTACCAACCAAGTTTC	130	
QY	808	TATCTGAACGGGAAACCTGTCTTTTGAAGGGCTTTGGAAGCAGAGAAATCCCCGTT	867	
DB	131	CTCATCAACAAAAGCCCTTCTACTTCCACGAGTAATTAACAGAGACTCTGATATT	190	
QY	868	CTGGGGCAGGGCACCCTTTTATCCATGTGATATAAAGACTTCAACTTCTGAAGTGATC	927	
DB	191	CGAGGCCAAGGCGCTGGAGCTGGCCCCCTCATGTGTGAAGGACTTTAATTATTGAAGTGGTTG	250	
QY	928	AAAGCGAATTCTTTCAGGACCTCTCACTATCTCTTACAGTGAAGAGTGGCTGATCTTGGC	987	
DB	251	GGGGCCCAACTCGTTCGACCAACCACTACCTTATGACAGAGAGATCCTGCAGATGTGT	310	
QY	988	GACAGACTCGGAATCCTTGTGATAGACGAAGCCCGCACGTTGGTATCACA--AGGTAC	1044	
DB	311	GACCGCCATGGCATCGTGTGTGATAGACGAGTGCCGGCGGTGGCATAAAGACATTCGC	370	
QY	1045	CACTACAATCCCGAGACTCAGAAAGATAGCAGAAGACAACATTAAGAAATGATGCACAGA	1104	
DB	371	AGTTTGGAAACGCTCTTAAACCATCACCTGTGCTCATGAGCAGAGCTGTACGTGG	430	
QY	1105	CACAAGAACCATCCAGTGTGATCATGTGAGTGGCGAAGCAACAGAGTCCAAACAT	1164	
DB	431	GACAAAGAACCATCCCTCTGTGTGATGTGTGTGAGGCCAATGAGCCGCTGCAGAGATG	490	
QY	1165	CCAGACGCGGAGGTTTCTTCAAAGCCCTTTATGAGACTGCCAATGAATGATGCAACA	1224	
DB	491	CCTCTGTGATTAATTATTTCAAACCTTGATTAACATACCAAGAATTGGATCCAACC	550	
QY	1225	CGCCCCGTTGTCAATGTGAGCATGATGACGACGACGACGAGAGAAAGAGAGCTGGCG	1284	
DB	551	CGGCCCCGT-----CACTTTTATCAGACAGAGTAAGTAATGACGAGATTAAGGG	598	
QY	1285	CTGAAGTACTTGCACATCGTCTGTGTGAACAGAGTACTACGCTGGTACATCTATCAGGGA	1344	
DB	599	GCTCCCTACGTGACGTCACTGTGGTAAACAGTTACTTCTCTGTGTAACATGACCCGGGC	658	
QY	1345	AGGATAGAAAGAGACTTCAAGCTCTGAAAAAGACATAGAAAGAGCTCTATGCAAGGCAC	1404	
DB	659	CACCCGAGGTCAATCCCATCCAGCTCAACACTCAGTTTGAGAACTGTACGGAAGTAC	718	
QY	1405	AGAAAGCCCATCTTGTACAGAAATTCGGTGGGACGCGATAGCTGGCATCCACTACGAT	1464	
DB	719	CAGAAACCATCATCCAGAGGAAATACGAGCGGATGCGGTGCGGGGCTTCAACAGTGAT	778	
QY	1465	CCACCTCAAAATGTTCTCCGAAGAGTACCAAGCAGAGCTGTTGAAAAAGATCAGGCTC	1524	
DB	779	CCACCCGTGATGTTTACTGAGGAGTACCAAGATTAGTCTGACAGACTACCAACAGCTG	838	
QY	1525	CTTTGAAAAAG-----ACTACATCATCGAACAACACGTTGGGCTTTGCAATTTTA	1579	
DB	839	TTCGACCAAGAAAGACAGTACGTCACTCGGGAATCATCTGGAATTTGCAAGCTTCA	898	
QY	1580	AGACTCCTCAGAAATGTGAGAAGCCATTCTCAACACAAAGGTTGTTTCAAGAGACA	1639	

DB	899	TGACCACACAAGGATCATGCG-CGTGTGGGGAACAGAGGTTGTTCTCAGACGAAAA	957	
QY	1640	GACAAACCCAAAC	1651	
DB	958	GGCAGCCCAAGC	969	

RESULT 6

AK041058

LOCUS

DEFINITION

AK041058 2274 bp mRNA linear HTC 03-APR-2004  
Mus musculus adult male aorta and vein cDNA, RIKEN full-length enriched library, clone:A530072005 product:beta-glucuronidase structural, full insert sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

AK041058.1 GI:26088231  
HTC; CAP trapper.  
Mus musculus (house mouse)

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W., Hayashida,K., Hayatsu,N., Hiramoto,K., Hirooka,T., Hirozane,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T., Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M., Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N., Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.  
Direct Submission  
Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of

3  
Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsuina,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.  
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer  
Genome Res. 10 (11), 1757-1771 (2000)

4  
The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.  
Functional annotation of a full-length mouse cDNA collection  
Nature 409, 685-690 (2001)

5  
The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.  
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs  
Nature 420, 563-573 (2002)



Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.jp, URL: <http://genome.gsc.riken.jp/>, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

cDNA library was prepared and sequenced in Mouse Genome  
 Encyclopedia Project of Genome Exploration Research Group in Riken  
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
 Division of Experimental Animal Research in Riken contributed to  
 prepare mouse tissues.  
 Please visit our web site for further details.  
 URL: <http://genome.gsc.riken.jp/>  
 URL: <http://fantom.gsc.riken.jp/>.

## FEATURES

```

source
1. .2274
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="PANTOM_DB:A530072005"
/db_xref="taxon:10090"
/clone="A530072005"
/sex="male"
/tissue_type="aorta and vein"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
14. .1778
/note="beta-glucuronidase structural (MGD|MGI:95874,
GB|NM_010368, evidence: BLASTN, 99%, match=733)
putative"
2260. .2265
/note="putative"
2274
/note="putative"

```

ORIGIN

Query Match	9.4%;	Score 159.6;	DB 3;	Length 2274;
Best Local Similarity	50.6%;	Pred. No. 1.7e-35;		
Matches 477;	Conservative	0;	Mismatches 444;	Indels 21;
				Gaps 31

[illegible]

Db	1239	GACCTCACCCGTCGCCGTGACCTTTGTGAGCA-----ACGCCAATAATGATGCA	1286
QY	1276	GACGTGGCGCTGAAGTACTTCGACATCGTCTGTGTGAACAGGTACTACGGCTGTACATC	1335
Db	1287	GACCTGGGGGCCCGCTACGTGAGATGTATCTGTGTAAACAGCTACTTTTCTTGGTATCAT	1346
QY	1336	TATCAGGGAAGATAGAGAAGAACTTCAAGCTCTGAAAAAGACATAGAGAAGCTCTAT	1395
Db	1347	GACTATGGGCATTTGGAGGTGATTCAGCCACAGCTGAATAGCCAGTTTGAGAACTGSTAT	1406
QY	1396	GCAAGCACACAGAAAGCCCATCTTTGTCAAGAAATTCCGTGCGGACCGCATGCTGGCATC	1455
Db	1407	AAGAGCGATCAGAAAGCCGATTAATCCAGAGCGAGTATGAGACAGACGCAATCCCAAGGATC	1466
QY	1456	CAC TACGATCCACCTCAATGTCTCCGAGAGTACCACAGCAGAGCTCGTTGAAAAAGACG	1515
Db	1467	CACGAGGACCCGCCCTCGCATGTTCAGTGAAGAGTACCAGAAAGCTGTTCTGGAGAAATTAC	1526
QY	1516	ATCAGGCTCCTTTTGAAA-----AAAGACTACATCATGTGGAAACACACGTTGGGCTTTT	1569
Db	1527	CATTCA GTTCTGGATCAGAAACGTAAAGAAATACGTGTCGGAGAGAGCTCATCTGGAATTTTC	1586
QY	1570	GCAGATTTTAAAGACTCCTCAAGATGTGAGAGAACCCATTTCTCAACCACAGAGGGTGTTTTC	1629
Db	1587	GCCGACTTTCATGACGAACCAAGTCAACCGCTGAGAGTAATTCGAAAACAAGAGGGATCTTC	1646
QY	1630	ACAAGAGACACAAACCCAAACTCGTTGCTCATGTACTGAGA	1671
Db	1647	ACTCGCCAGAGACAGCCCAAACTTCGGCCTTTATTTTTCGA	1688

RESULT 7  
CK467358

LOCUS	CK467358	816 bp	mRNA	linear	EST 14-JAN-2004
DEFINITION	938649 MARC 4PIG Sus scrofa cDNA 5', mRNA sequence.				
ACCESSION	CK467358				
VERSION	CK467358.1	GI:40838635			
KEYWORDS	EST.				
SOURCE	Sus scrofa (pig)				
ORGANISM	Sus scrofa				

## REFERENCE AUTHORS

TITLE	Nonnemman,D.J., Wray,J.E. and Keele,J.W. Porcine EST collection using a normalized library constructed from embryos representing early developmental stages Unpublished (2003)
JOURNAL	
COMMENT	Contact: Smith TPL

JOURNAL  
COMMENT

USDA, ARS, US Meat Animal Research Center  
PO Box 166, Clay Center, NE 68933-0166, USA  
Tel: 402 762 4366  
Fax: 402 762 4390  
Email: smith@email.marc.usda.gov  
Single pass sequencing. Bases called with phred v0.020425.c and  
trimmed with the aid of the trim\_alt option. Vector identified with  
cross\_match v0.990329.  
Plate: TMM8042 row: M column: 18  
Seq primer: GTAAATACGACTCACTATAGGG.

## FEATURES

```

1. .816
/organism="Sus scrofa"
/mol_type="mRNA"
/db_xref="taxon:9823"
/tissue_type="pooled"
/lab_host="DH10B"
/clone_lib="MARC 4PIG"
/notes="Vector: pCDNA3.1; Site_1: EcoRI; Site_2: NotI"
Library made with combined RNA from day-10, day-13,
day-15, day-25, and day-30 whole embryos."

```

ORIGIN

Query Match		8.9%;	Score 150.6;	DB 7;	Length 816;
Best Local Similarity		53.2%;	Pred. No. 5.8e-33;		
Matches 396;		Conservative 0;	Mismatches 334;	Indels 15;	Gaps 3;
QY	751 GACGAGTACACTCTGGACATCGGAATCAGAACGATCAGCTGGGACGAGAAAGGCTCTAT	810			
DB	84 GACTTCTACACCTCCCTGTGGGATTCGACACCTGGCTGTCAACAGAGACAGTTCCTC	143			
QY	811 CTGAACGGGAAACCTGTCTTTTGAAGGGCTTTGGAAGACGACGAGAAATCCCGTTCTG	870			
DB	144 ATCAACGGGAAACCTTCTATTTCCATGGGGTTAACAAGCAGAGATGACATATCCGA	203			
QY	871 GGGCAGGGCACCTTTTATCCATGATGATAAAGACTTCAACCTTCTGAAGTGATCAAC	930			
DB	204 GGCAGGGGCTTTGACTGTGCTGCTGTGGTGAAGGACTTCAACCTGCTCCGCTGGCTGGT	263			
QY	931 GCGAATTCCTTCAGGACCTCTCTCATCTATCCTTACAGTGAAGAGTGCTGATCTGGCCGAC	990			
DB	264 GCCAACGCTTCCGACACGACCACTACCCCTACGACAGAGAGTGATGACGCTGTGAC	323			
QY	991 AGACTCGGAATCCTTGTGATAGACGAAGCCCGCACGTTGGTATCACAAAGTACCA---C	1047			
DB	324 CGTTATGGGATCGTGTGATCATGATAGAGATCCGGGTGTGGGCACTGTGCTGGCCAGAGC	383			
QY	1048 TACAATCCCGAGACTCAGAAATAGCAGAAAGACAACTAAGAGAATGATCGACAGACAC	1107			
DB	384 TTCAGCAACGCGTCTCTGACAGACCACTAGAGGTGATGAGAGATGTTCCAGAGGAC	443			
QY	1108 AAGAACCATCCAGTGTGATCATGTGAGTGTGGGAAAGAACAGATCCAACTCA	1167			
DB	444 AAGAACCAACCCGCGTGTGATGTGTCGGGCCAATGAGCCTTCTTCTCTGGAA	503			
QY	1168 GACGCGGAGGTTCTTCAAGCCCTTTATGAGACTGCCAATGAATGATCGAACACGC	1227			
DB	504 CAGGCTGCTTACTTCAAGATGTGATCGGCCACACCAAGCCTTGACCCCTCCCG	563			
QY	1228 CCCGTTGTCATGTGAGCATGATGACGACGACGACGAGAGACAAGACGTCGCTG	1287			
DB	564 CCCGTGACCTTCGTGACCA-----GCTCAGCTATGA----AAAGACCTGGGGGTG	611			
QY	1288 AAGTACTTGCATCTGTGTGAAACAGTACTACGGCTGTATCATCTATCAGGGAAG	1347			
DB	612 CCGTATGTGACGTGATCTGTGTGAACAGTACTCTCTGTATCATGACTACGGTAC	671			
QY	1348 ATAGAAGAAGACTTCAAGCTCTGGAAGAAAGACATAGAAGACTCTATGCAAGCACAGA	1407			
DB	672 ATGAGAGTGATTACGCTGCAGCTGGCCACCCAGTTTGAGAGTGGCATGAGGCTTACAG	731			
QY	1408 AAGCCCATCTTTGTCAAGAAATTCGGTGGGACCGGATAGCTGCATCCACTACGATCCA	1467			
DB	732 AAACCAATTTATCCAGAGCGAGTACGGGGCAGAAACCATCATAGGGTTTCAACGAGATCCA	791			
QY	1468 CCTCAATGTTCTCCGAAGATACC	1492			
DB	792 CCTCTCATGTTCAAGTAAGATAACC	816			
RESULT 8					
LOCUS	BO941196	906 bp	mRNA	linear	EST 21-AUG-2002
DEFINITION	AGNCOURT_8741562 NIH_MGC_18 Homo sapiens cdna clone IMAGE:6420519				
ACCESSION	BO941196				
VERSION	BO941196.1 GI:22356674				
KEYWORDS	EST.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
AUTHORS	NIH-MGC http://mgc.ncl.nih.gov/.				
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)				
JOURNAL	Unpublished (1999)				

COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgapbs-remail.nih.gov Tissue Procurement: DCTD/DTP/Gazdar CDNA Library Preparation: Rubin Laboratory CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LINL) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LINL at: <a href="http://image.lnl.gov">http://image.lnl.gov</a> Plate: LCM2594 row: g column: 16 High quality sequence stop: 765.				
FEATURES	location/Qualifiers				
source	1. 906 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="IMAGE:6420519" /tissue_type="large cell carcinoma" /lab_host="DH10B (phage-resistant)" /clone_lib="NIH_MGC_18" /note="Organ: lung; Vector: pOTB7; Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCAAGAG(G). Library constructed by ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT. (Life Technologies). Note: this is a NIH_MGC Library."				
ORIGIN					
Query Match	8.5%;	Score 144.4;	DB 5;	Length 906;	
Best Local Similarity	52.1%;	Pred. No. 4.3e-31;			
Matches 454;	Conservative 0;	Mismatches 396;	Indels 22;	Gaps 5;	
QY	828 CTTTTGAAGGCTTTGAAGCAGAGAAATCCCGTCTGGGGCAGGCACTTTTA	887			
DB	1 CTATTTCCAGGTGTCAACAGCATGAGATCGGACATCGAGGGAAGGCTTCGACTG	60			
QY	888 TCCATTGATGATAAAGACTTCAACCTTCTGAAGTGATCAACGGAATCTTTCAGAC	947			
DB	61 GCCGCTGCTGTGAAGACTTCAACCTGCTGCTGGTGGTCCAAAGCTTCCGTAC	120			
QY	948 CTCTCACTATCTTACAGTGAAGAGTGGCTGATCTTGGCCAGACTCGGAATCCTTGT	1007			
DB	121 CAGCCACTACCCCTATGCAAGAGAGATGATGATGTGTGACCCGTATGGGATTTGTGT	180			
QY	1008 GATGACGAAGCCCCGACGTTGGTATCAACAAGTACCA---CTACAATCCGAGACTCA	1064			
DB	181 CATCGATGATGTCCCGCGTGGGCCCTGGCGCTGCCGAGTTCTTCAACAACGTTTCTCT	240			
QY	1065 GAAAGTAGCAGAAGACAACATTAAGAAGATGATCGACAGACACAAGAACCATCCAGTGT	1124			
DB	241 GCATCACCAATGCAAGGTGATGAAGAAGTGTGCTAGGGACAAGAACCCCGCGGT	300			
QY	1125 GATCATGTGAGTGTGGCGAAGCAACGAGTCCAAACCATCCAGACGGAGGTTTCTT	1184			
DB	301 CGTGATGTGTGTGTGGCCAAAGCCTGCGTCCACCTAGAATCTGCTGCTACTACTT	360			
QY	1185 CAAAGCCCTTTATGAGACTGCCAATGAATGATCGAACACGCCCCGTTGTATGTGTAG	1244			
DB	361 GAAAGTGTGATGCTCACACCAAAATCCTTGAACCCCTCCGCGCTGTGACCTTTGTGAG	420			
QY	1245 CATGATGAGCGCACCGAGCAGAGAACAAGAGAGCTGGCGCTGAAGTACTTGCACATCGT	1304			
DB	421 CAACTTAATATGACG-----CAGACAAGGGGGCTCCGTATGTGATGTGAT	468			
QY	1305 CTGTGTGAACAGTACTACGGCTGTATCATCTATCAGGGAAGGATGAAGAAGCACTTCA	1364			
DB	469 CTGTTGAACAGCTACTACTTGTGTATCAAGACTACGGGCACTGGAGTTGATTCAGCT	528			
QY	1365 AGCTCTGAAAAAGACATAGAAGAGCTCTATGCAAGGACAGAAAGCCCATCTTGTCTAC	1424			
DB	529 GCAGCTGGCCACCAGTTTGAAGAACTGTATTAAGAAGTATCAGAAGCCCATTAATTCAGAG	588			

RESULT 9				
CN155220				
LOCUS	CN155220	878 bp	mRNA	linear
DEFINITION	942501 MARC 4PTG Sub scrofa cdna 5', mRNA sequence.			EST 02-APR-2004

REFERENCE  
1 (bases 1 to 878)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.

AUTHORS	Smith, T.P.L., Frickling, B.A., Ford, J.J., Vallet, J.L., Wise, T.A., Nonneman, D.J., Wray, J.E. and Keele, J.W.
TITLE	Porcine EST collection using a normalized library constructed from embryos representing early developmental stages
JOURNAL	Unpublished (2003)
COMMENT	Contact: Smith TPL

JOURNAL  
COMMENT

Unpublished (2003)  
Contact: Smith TPL  
USDA, ARS, US Meat Animal Research Center  
PO Box 166, Clay Center, NE 68933-0166, USA  
Tel: 402 762 4366  
Fax: 402 762 4390  
Email: smith@email.marc.usda.gov  
Single pass sequencing. Bases called with phred v0.020425.c and  
trimmed with the aid of the trim\_alt option. Vector identified with  
cross\_match v0.990329.  
Plate: TMM8046 row: I column: 19  
Seq primer: GTAATACGACTCACTATAGG.

FEATURES	Location/Qualifiers
source	1. .878

```

/organism="Sus scrofa"
/mol_type="mrna"
/db_xref="taxon:9823"
/tissue_type="pooled"
/lab_host="DH10B"
/clone_lib="MARC 4PIG".
/note="Vector: pCDNA3.1; Site_1: EcoRI; Site_2: NotI;
Library made with combined RNA from day-10, day-13,
day-15, day-25, and day-30 whole embryos."

```

**ORIGIN**

Query Match	8.4%;	Score 142.6;	DB 7;	Length 878;
Best Local Similarity	52.1%;	Pred. No. 1.5e-30;		
Matches 425; Conservative	0;	Mismatches 369;	Indels 21;	Gaps 4;

Oy		866	TTCCTGGGCGAGGCACCTTTTATCCATTGATGATAAAGA CTTCAACCTTCGAAGTGA	925
Db		6	TCCGAGGCAAGGGCTTTGACTGTGCTGCTGTGTAAGACTTCAACCTGCTCCGCTGGC	65

926 TCACGGCAATCTTTCAGGACCTCACTATCCTTACAGTAGAGTGGCTGATCTTG 985

Db	66	TGGGTTGCCAAGC	CTTCCGCA	CCAGCC	ACTA	CCCCCT	ACGCA	GAGG	AGTG	ATG	CA	GC	TCT	125								
QY	986	CCGAC	CAGACT	TGGGA	ATC	CTTGT	GAT	TAG	ACGA	AGCCCC	CGC	AGC	TTG	TAT	CACA	AGGT	ACC	1045				
Db	126	GTG	ACCG	TTAT	TGGG	ATCG	TGT	CAT	CGAT	GAGA	GTCC	CGGT	GTGG	GCAT	CGT	GTG	CCCC	185				
QY	1046	A---	CTACA	ATCCC	CGAG	CTC	GAGA	GAT	AGAGA	CAACA	CAT	TA	GMA	GAT	G	TC	GACA	1102				
Db	186	AGAG	CTT	CAGCA	ACGCG	TCTCT	GTG	CAG	CACCA	CC	TAG	AGG	TGAT	GAG	GAG	ATG	TTG	CA	245			
QY	1103	GAC	ACA	GAA	GAAC	CA	TCC	CAG	TGT	GAT	CA	TGT	GSA	GTG	GG	CGA	ACGA	ACC	1162			
Db	246	GGG	ACA	GAA	GAAC	CA	CCCG	GCG	CTT	GTG	ATG	TGT	GTCC	GCG	CA	ATG	AG	CC	CTT	CC	TTC	305
QY	1163	A	TCC	AGA	CGCG	GAG	GGT	TTCT	TCA	AAG	CCCT	T	TAT	GAG	ACT	GCC	AATG	AAAT	TG	AT	CGAA	1222
Db	306	TGG	AACA	GAG	CTG	CTT	ACT	ACT	TTCA	AGAT	GT	CG	GC	ACACA	CCAA	AG	CC	TTG	GAA	CCCC	CT	365
QY	1223	CAC	GCCCC	CGT	TG	TCAT	GTG	TAG	CA	TGAT	GAC	GC	CAC	CAG	ACG	AGA	GAA	CAAG	AG	AC	GTGG	1282
Db	366	CCC	GCCCC	CGT	GA	CTT	CG	TG	ACCA	-----	GCT	CC	AG	CT	AT	G	-----	AAA	AG	AC	CTGG	413

OY		1283	CGCTGAAGTACTTCGACATCGTCTGTGTGAACAGGTACTACGGCTGTACTATTACAGG	1342
Db		414	GGGTGCCGTAITGTGACCGTATCTGTGTGAACAGTTACTACTCCTGTGATCATGACTACG	473
OY		1343	GAAGGATAGAAGAAAGGACTTCAAGCTCTGSAAAAAAGACATAGAAGAGCTTATGC AAGGC	1402
Db		474	GTCACATGGAGGTGATTTCAGCTGCAGCTGGCCACCAGTTTGAAGAGGTGGCATGAGGCTT	533
OY		1403	ACAGAAAGCCCATCTTTGTTCACAGAATTCGGTGGGACGGGATAGCTGGCATCCACTACG	1462
Db		534	ACCAGAAACCAATTATCCAGAGCGAGTACGGGGCAGAAACCATCATAGGCTTTCACGAGG	593
OY		1463	ATCCACCTCAATGTCTCCGAAGAGTACCAAGCAGAGCTCGTTGAAAAGACGATCAGGC	1522
Db		594	ATCCACCTCATGTTCAGTGAAGAATACCAAGAAAGGCGCTGTACAGCACGTATCATGTGA	653
OY		1523	TCCTTTTGAAA-----AAAGACTACATCATCGGAACAACAGCTGTGGCCCTTGCAGATT	1576
Db		654	TTCCTGATCAAAAAAGCAAAGATATGTGGTTGGAGAGCTCATCTGAAATTTTGGCGATT	713
OY		1577	TTAAGACTCCTCAGAATGTGAGAAGACCATTCTCAACCACAAGGGTGTTCACAAGAG	1636
Db		714	TTATAGCCGACCAATCACCTCAGAGGGCCATAGGAATAAGAAAGGATCTTCACTCGAC	773
OY		1637	ACAGAACCAACCAACTCGTTGCTCATGTACTGAGA	1671
Db		774	AGAGACAACCAAAAAAGTGCAGCATTCCTGTGGCA	808

## RESULT 10

LOCUS	1051 bp	mRNA	linear	EST 08-APR-2004
DEFINITION	BX363460	Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED		
	BX363460	Homo sapiens cDNA clone CS0DL001YM21 5-PRIME, mRNA sequence.		

ACCESSION	BX363460
VERSION	BX363460.2
	GI:46291531

KEYWORDS	EST.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1	Li, W.B., Gruber, C., Jessee, J. and Polayes, D.	Full-length cDNA libraries and normalization	Unpublished (2001)	On May 5, 2003 this sequence version replaced q1:30372720.

Genoscope - Centre National de Séquençage  
BP 191 91006 Evry cedex - France  
Email: [segref@genoscope.cns.fr](mailto:segref@genoscope.cns.fr), Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr)  
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime







ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 900)  
NIH-MGC <http://mgs.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: [csapbs-remail.nih.gov](mailto:csapbs-remail.nih.gov)  
Tissue Procurement: Kristi A. Egland, Ira Pastan  
CDNA Library Preparation: Invitrogen Corp  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.lnl.gov>  
Plate: LLM14283 row: d column: 20  
High quality sequence stop: 580.

FEATURES  
source  
1.900  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6721700"  
/cell\_line="ZR-75-1, MCF7, SK-BR-3, MDA-MB-231, hTERT-HME1, LNCap"  
/lab\_host="EMDH10B"  
/clone\_lhb="MAPCL"  
/note="Vector: PCMV-SPT6; Site 1: EcorV; Site 2: Not I; Subtracted with brain, liver, lung, kidney and muscle. Directionally cloned. Priming method: oligo-dT. Average insert size: 1800 bp. Library amplification: 26,000 fold. Kristi A. Egland, James J. Vincent, Robert Strausberg, Bungkok Lee & Ira Pastan: Discovery of new breast cancer genes encoding membrane and secreted proteins. Manuscript submitted."

ORIGIN

Query Match	7.7%;	Score 129.8;	DB 6;	Length 900;
Best Local Similarity	51.9%;	Pred. No. 9.8e-27;		
Matches 402;	Conservative 0;	Mismatches 352;	Indels 21;	Gaps 4;
866	TTCTGGGGCAGGGC	CTTTTATCATTTGATGATTA	AAAGACTTCAACCTTCTGAAGTGA	925
31	TCCGAGGGAGGGCTT	CGACTGGCCGCTGCTGTGAAGACTTCA	ACTGCTTCGCTGGC	90
926	TCAACGGCAATTTT	CAGACCTCTCACTATCTTACAGTGAAG	TGGCTGATCTTG	985
91	TTGTGTCCAACGCTT	TCCGTACAGCACCTACCCCTATG	CAGAGAAAGTATGATGATGT	150
986	CCGACAGACTCGGA	ATCCTTGTGTATGACGAAGCCCCG	CACGTTGGTATCAAGGTACC	1045
151	GTGACCGCTATGG	ATTTGTGTATGATGATGTCCGCGCTG	CGCCGCTGCGC	210
1046	A---CTACAATCCG	AGACTCAGAGATAGCAGAGACAACA	CATTAAGAAGATGATGACA	1102
211	AGTTCTTCAACA	CGTTTCTCTGCATCACCACATG	CAGGTATGGAAGAAGTGTGCGTA	270
1103	GACACAAGAACCA	TCCAGTGTGATCATGTGTGGCGGA	AGAACAGACAGTCCAAAC	1162
271	GGGACAAGAACCA	CCCCGCGTGTGTGTGTGGCCAA	GAGACCTGCTGCCACC	330
1163	ATCCAGACGCGG	AGGTTTCTTCAAAAGCCCTTAT	GAGACTGCCAATGAATGATCGAA	1222
331	TAGAAATCTGCTG	CTACTACTTGAAGATGTGTATGCT	CACACCAATCTTGAACCCCT	390
1223	CACGCCCCGTTG	TCATGTGTAGCATGATGACGCA	CCAGACGAGAGAACAAGACGTGG	1282
391	CCCGGCTGTG	ACTTTGTGAGCACTTAATATG	CAG-----CAGACAAGG	438
1283	CGCTGAAGTACT	TGCATCGTCTGTGTGAACAG	GTACTACGCTGTATCATATACAG	1342
439	GGGCTCCGATG	TGATGTGATCTGTTGAACG	CTACTACTTGTGTATCAGACTACG	498

Query Match	7.7%	Score 129.8	DB 6	Length 914
Best local Similarity	50.4%	Pred. No. 9.8e-27		
Matches 410	Conservative 0	Mismatches 382	Indels 21	Gaps 3

```

QY      871 GGGCAGGCGACCTTTATCCATGTGATATAAAGACTTCAACTTCTGAAAGTGATCAAC 930
Db      3 GGGAAAGGCTTCGACTGGCCGCTGTGTAAAGATTCAACTGCTCCGTTGGCTCGG 62
QY      931 GCGAATTCCTTCAGGACCTCTCACTATCTCTTAAGTGAAGAGTGCTGATCTTGCCGAC 990
Db      63 GCAATTCCTTCGTACCAAGCACTATCTCTACTCAGAGAGGTAAGTCTGAGCTGTGAC 122
QY      991 AGACTCGGAATCCTGTGTGATAGACGAAGCCCGCACGTTGG--TATCACAAGGTACCAC 1047
Db      123 CGATACGGGATTTGTGTATCATGAGTGTCCGGTGTGGGCAATTTGCTAAGTCTCAGAGT 182
QY      1048 TACAATCCCGAGACTCAGAGATAGCAGAGACACAATAGAAGATGATCGACAGACAC 1107
Db      183 TTTGGCAACGAGTCACTTCGGCAACCACTAGAGGTGATGAGAGAGCTGTTCCGGGAC 242
QY      1108 AAGAACCATCCAGTGTGATCATGTGAGTGTGGCAACGACCAAGAGTCCAAACCATCCA 1167
Db      243 AAAAATCACCCCTGGCGTTGTGATGTGTGTGTGGCCAATGAGCCTTCTCTGTCTGAAA 302
QY      1168 GACGCGGAGGGTTTCTTCAAAAGCCCTTTATGAGACTGCCAATGAATGATCGAACACGC 1227
Db      303 CCCGCCGATATTACTTTAAGACGCTGATCAACCAACCAAGCCCTGAGCTCACCCTG 362
QY      1228 CCCGTTGTATGTGATGATGAGACGACCAACGAGAGAACAGAGAGAGAGAGAGAGAGAG 1287
Db      363 CCCGTGACCTTTGTGAGCA-----ACGCCAAATATGATGACAGACCTGGGGGCC 410
QY      1288 AAGTACTTCGACATCGTGTGTGTGACAGAGTACTACGGCTGTACATCTATCAGGGAAG 1347
Db      411 CCGTACGIGATGTTATCTGTGTAAACAGTACTTTCTTGTGTATCATGACTATGAGGCAT 470
QY      1348 ATAGAAGAAAGACTTCAAGCTCTGAAAAAGACATAGAAGAGCTTATGCAAGGACACAGA 1407
Db      471 TTGAGGTGATTCAGCCACAGCTGAATAGCCAGTTGAGAACTGTATAGAAGACGATCAG 530
QY      1408 AAGCCCATCTTTGTACACAGATTGCGTGCGAGCGGATAGCTGGCATCCACTACGATCCA 1467
Db      531 AAGCCGATTATCCAGAGCGAGTATGAGAGCAGCAATCCCAAGGATCCAGAGGACCCG 590
QY      1468 CCTCAATGTCTCCGAGAGTACCAAGCAGAGCTCGTTGAAA-----AGACGATCAGG 1521
Db      591 CCTCGCATGTTCAAGTAGAGAGTACCAAGAGGCTGTTCTGAGAAATTACCATTCAGTTCTG 650
QY      1522 CTCCTTTGAAAAAAGACTATCATCATCGGAACACACAGTGTGGGCTTTGCGAGATTTAAG 1581
Db      651 GATCAGAAACGTAAAGATACGTGTGCGAGAGCTCATCTGGAATTTCCGCGACTTCATG 710
QY      1582 ACTCCTCAGATGTGAGAAAGACCCATTCTCAACCAAGAGGTGTTTCAAGAGAGACAGA 1641
Db      711 ACGAACCAAGTCAACCGCTGAGAGTAAATCGGAAACAAGAGGGGATCTTCACTCGCCAGAGA 770
QY      1642 CAACCCAAACTCGTGTCTCATGTACTGAGAGA 1674
Db      771 CAGCCCAAACTTCGGCCTTTATTTTGGGAGA 803

```

```

RESULT 15
LOCUS      CO394571      862 bp      mRNA      linear      EST 01-JUL-2004
DEFINITION AGENCOURT 27822917 NIH MGC_252 Rattus norvegicus cDNA clone
IMAGE:7307606 5', mRNA sequence.
ACCESSION  CO394571
VERSION     CO394571.1 GI:49576487
KEYWORDS    EST.
SOURCE      Rattus norvegicus (Norway rat)
ORGANISM    Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1 (bases 1 to 862)
AUTHORS    NIH-MGC http://mgi.nci.nih.gov/.
TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)

```

## JOURNAL COMMENT

Unpublished (1999)  
 Contact: Daniela S. Gerhard, Ph.D.  
 Office of Cancer Genomics  
 National Cancer Institute / NIH  
 Bldg. 31 Rm10A07 Bethesda, MD 20892  
 Email: c9apbs-r@mail.nih.gov  
 Tissue Procurement: Drs. Josef Lazar & Howard Jacob, Medical  
 College of Wisconsin  
 cDNA Library Preparation: Express Genomics  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
 http://image.lnl.gov  
 Plate: L1AM15349 row: e column: 12  
 High quality sequence stop: 685.

## FEATURES

## source

```

1..862
/organism="Rattus norvegicus"
/mol_type="mRNA"
/db_xref="taxon:10116"
/clone="IMAGE:7307606"
/tissue_type="Ovary, strain - Norway Line3. Age8 weeks.
Tissue was snap-frozen adn transferred in -70C. RNase free
for the entire procedure"
/lab_host="DH10B Tona"
/clone_1ib="NIH MGC_252"
/notes="Organ: ovary; Vector: pDONR 201; Site_1: NotI;
Site_2: NotI; RNA obtained from female ovaries animals at
8 wk old. Tissues were snap-frozen and kept at -80C for
two days before RNA extraction and purification
(Tri-reagent method). cDNA was primed using oligo-dT
primer: 5'-pGACTAGTCTTAGATCGGAGCGGCCGCC(T)25-3' and
cloned into the EcoRV/NotI sites of pExpress-1.
Size-selection >1.25kb resulted in an average insert size
of 1.7kb. This primary library is not normalized
(normalized library is NIH MGC 252) and was constructed by
Express Genomics (Frederick, MD). Note: this is a NIH-MGC
library"

```

## ORIGIN

```

Query Match      7 6%; Score 128.2; DB 7; Length 862;
Best Local Similarity 51.5%; Pred. No. 2.9e-26;
Matches 354; Conservative 0; Mismatches 318; Indels 15; Gaps 2;

QY      855 GGAATTCCTCCGTTGCGGAGGAGCACTTTATCCATGTGATATAAAGACTTCAACT 914
Db      7 GGATTCCTCCGATCGAGGAGAGGCTTCAGCTGCTCTGCTGATTAAGATTCAACT 66
QY      915 TCTGAAGTGATCAACGCGAATCTTTCAGAGCTCTCACTATCTTACAGTGAAGAGTG 974
Db      67 CTTCCGTTGGCTCGGGCAAAATCTCTTCTGACCAAGCACTATCTCTACTCGAGAGAGT 126
QY      975 GCTGATCTTGCCGACAGACTCGGAATCTTGTGATAGACGAAGCCCGACAGTGTAT 1034
Db      127 ACTTCAGCTCTGTGACCGATATGATGTGTGATGATGAGTGTCCGCTGTGGCAT 186
QY      1035 CACAGGTACCA---CTACAATCCCGAGACTCAGAAAGATAGCAGAAAGACACATTAAGAG 1091
Db      187 CGTGTGCCCCAGAGTTTGGCAACGTGTCTCTTCGGCACCACTAGAGAGTGAAGCA 246
QY      1092 AATGATCGACAGACACAAGAACCATCCAGTGTGATCATGTGAGTGTGGCAAGAAC 1151
Db      247 GCTGTGCGCAGAGGACAAATCACTCGCGGTGTGATGTGTGTGTGGCCAATGAGCC 306
QY      1152 AGAGTCCAACCATCCAGACGCGGAGGTTTCTTCAAAAGCCCTTTATGAGACTGCCAATGA 1211
Db      307 TGTCTCTTCTCTGAACCTGCGGATATTACTTCAAGACGCTGATGCGCCACACCAAGC 366
QY      1212 AATGATCGAACAGACGCCCGTTGTGATGTGAGCATGATGAGCGCACAGAGAGAGAAC 1271
Db      367 CTTGAGCCCAACCCGCTCTGTGACTTTGTGAGCAATA-----CCAGATATGA 414

```

```
QY      1272 AAGAGACGTGGCCGTGAAGTACTTCGACATCGTCTGTGTGAACAGGTACTACGGCTGTA 1331
      |||||
      415 CGCAGACCTGGGGGCCCCCGTACGTGACGTGATTTGTGAACAGTTACTTATCCTTGTA 474
      |||||
QY      1332 CATCTATCAGGGAAGGATAGAAGAAGACTTCAAGCTCTGGAAGAAAGACATAGAAGAGCT 1391
      |||||
      475 TCATGACTACGGGCATCTGGAGGTGATTCAGCTGCAGCTGACTAGCCAGTTTGAGAACTG 534
      |||||
QY      1392 CTATGCAAGGCACAGAAAGCCCATCTTTGTCAACAGATTCCGTCGGACGCGATAGCTGG 1451
      |||||
      535 GTATTAAGATGTACCAGAAAGCCCAATTATCCAGAGCGAGTATGGAGCAGACGCCGCTCGGG 594
      |||||
QY      1452 CATCCACTACGATCCACCCTCAATGTCTCCGAAGAGTACCAGCAGAGCTCGTTGAAA 1511
      |||||
      595 GCTTCATGAGGATCCACCCTCGCATGTTCAAGTGAAGGTACCAGACAGCTCTCTGGAGAA 654
      |||||
QY      1512 GACGATCAGGCTCCTTTGAAAAAGA 1538
      |||||
      655 TATCATTTGATCTGGATGAGAAAGA 681
      |||||
```

Search completed: January 24, 2005, 07:30:31  
Job time : 5577 secs

***This Page Blank (uspio)***

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 24, 2005, 05:57:45 ; Search time 100 Seconds  
(without alignments)  
3239.359 Million cell updates/sec

Title: US-09-936-759-6  
Perfect score: 3001  
Sequence: 1 MVRPQRNKKRFFILILNGVWN.....TRDRQPKLVAVHLRLWSEV 563

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : UniProt\_02:\*  
1: uniprot\_sprot:\*  
2: uniprot\_trembl:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2999	99.9	563	2 Q9X0F2	Q9X0F2 thermotoga
2	1011	33.7	570	2 Q97U11	Q97U11 sulfolobus
3	977.5	32.6	599	2 Q8XP19	Q8XP19 clostridium
4	972.5	32.4	599	2 Q8VNV4	Q8VNV4 clostridium
5	929	31.0	602	2 Q9AFA2	Q9AFA2 staphylococ
6	919.5	30.6	598	2 Q9AHJ8	Q9AHJ8 lactobacill
7	906	30.2	670	2 Q6NL66	Q6NL66 drosophila
8	906	30.2	670	2 AAS93746	AAS93746 drosophil
9	905	30.2	656	2 Q9V8R0	Q9V8R0 drosophila
10	905	30.2	670	2 Q8MMB7	Q8MMB7 drosophila
11	902.5	30.1	648	1 BGLR MOUSE	BGLR MOUSE
12	899.5	30.0	648	2 Q6IRI0	Q6IRI0 mus musculu
13	898	29.9	651	1 BGLR FELCA	BGLR FELCA
14	894.5	29.8	648	1 BGLR RAT	BGLR RAT
15	888	29.6	651	1 BGLR CANFA	BGLR CANFA
16	882	29.4	603	2 Q93VY4	Q93VY4 arabidopsis
17	880	29.3	603	2 Q8FHA3	Q8FHA3 escherichia
18	879	29.3	603	1 BGLR ECOLI	BGLR ECOLI
19	879	29.3	603	2 Q9AHJ4	Q9AHJ4 escherichia
20	872.5	29.1	648	1 BGLR CERAE	BGLR CERAE
21	865.5	28.8	651	1 BGLR HUMAN	BGLR HUMAN
22	865.5	28.8	651	2 AAQ96851	AAQ96851 homo sapi
23	862	28.7	628	2 Q95Q32	Q95Q32 caenorhabdi
24	861.5	28.7	603	2 Q6WJ77	Q6WJ77 ruminococcu
25	861.5	28.7	603	2 AAQ76046	AAQ76046 ruminococ
26	856.5	28.5	808	2 Q7TPJ3	Q7TPJ3 rattus norv
27	841.5	28.0	686	2 Q9V9T9	Q9V9T9 drosophila
28	826.5	27.5	599	2 Q8E6A6	Q8E6A6 streptococc
29	821.5	27.4	599	2 Q8E0N2	Q8E0N2 streptococc
30	811	27.0	567	2 Q7UCE6	Q7UCE6 shigella fl
31	809	27.0	567	2 Q83RCS	Q83RCS shigella fl

32	728	24.3	660	2 Q7PZE2	Q7PZE2 anopheles g
33	691	23.0	368	2 Q8X671	Q8X671 escherichia
34	691	23.0	370	2 Q7ADL5	Q7ADL5 escherichia
35	531	17.7	459	2 Q8FMX0	Q8FMX0 corynebacte
36	487	16.2	755	2 Q92XF7	Q92XF7 rhizobium m
37	486.5	16.2	745	2 Q93IM0	Q93IM0 thermoanaer
38	482	16.1	716	1 BGL_THETU	P26257 thermoanaer
39	482	16.1	716	2 CAA38462	CAa38462 thermoana
40	481.5	16.0	998	1 BGL_LACIA	Q48727 lactococcus
41	477.5	15.9	998	2 Q8VPU3	Q8VPU3 lactococcus
42	476.5	15.9	1085	2 Q85250	Q85250 thermotoga
43	472.5	15.7	996	2 Q87523	Q87523 lactococcus
44	460	15.3	743	1 BGL_THET	P77989 thermoanaer
45	452.5	15.1	1084	1 BGL_THEMA	Q56307 thermotoga

ALIGNMENTS

RESULT 1					
Q9X0F2	PRELIMINARY;	PRT;	563 AA.		
ID	Q9X0F2				
AC	Q9X0F2;				
DT	01-NOV-1999 (Tremblrel. 12, Created)				
DT	01-NOV-1999 (Tremblrel. 12, Last sequence update)				
DT	01-OCT-2003 (Tremblrel. 25, Last annotation update)				
DE	Beta-glucuronidase.				
GN	OrderedLocusNames=TM1062;				
OS	Thermotoga maritima.				
OC	Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.				
OX	NCBI_TaxID=2336;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=MSB8 / DSM 3109 / ATCC 43589;				
RX	MEDLINE=99287316; PubMed=10360571; DOI=10.1038/20601;				
RA	Nelson K.E., Clayton R.A., Gill S.R., Gwin M.L., Dodson R.J.,				
RA	Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,				
RA	McDonald L.A., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,				
RA	Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.L.,				
RA	Heidelberg J.F., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,				
RA	Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;				
RT	"Evidence for lateral gene transfer between Archaea and Bacteria from				
RT	genome sequence of Thermotoga maritima."				
RL	Nature 399:323-329(1999).				
DR	EMBL; AE001766; AAD36143.1; -				
DR	PIR; A72300; A72300.				
DR	HSSP; P08236; 1BHG.				
DR	TIGR; TM1062; -				
DR	GO; GO:0004553; F:hydrolyase activity, hydrolyzing O-glycosyl . . .; IEA.				
DR	GO; GO:0005975; P:carbohydrate metabolism; IEA.				
DR	InterPro; IPR008979; Gal_bind_like.				
DR	InterPro; IPR006101; Glyco_hydro_2.				
DR	InterPro; IPR006102; Glyco_hydro_2ig.				
DR	InterPro; IPR006104; Glyco_hydro_2SB.				
DR	InterPro; IPR006103; Glyco_hydro_2TIM.				
DR	Pfam; PF00703; Glyco_hydro_2; 1.				
DR	Pfam; PF02836; Glyco_hydro_2_C; 1.				
DR	Pfam; PF02837; Glyco_hydro_2_N; 1.				
DR	PRINTS; PR00132; GLHYDRLASE2.				
DR	PROSITE; PS00719; GLYCOSYL_HYDROL_F2_1; 1.				
KW	Complete proteome.				
SQ	SEQUENCE 563 AA; 65682 MW; 98C030B75D33B6C1 CRC64;				
Query Match					
Best Local Similarity 99.8%; Score 2999; DB 2; Length 563;					
Matches 562; Conservative 0; Mismatches 1; Indels 0; Gaps 0;					
Qy	1 MVRPQRNKKRFFILILNGVWNLEVTSKDRPIAVPGSWNEQYODLCYEEGPFYKTFYVPK 60				
Db	1 MVRPQRNKKRFFILILNGVWNLEVTSKDRPIAVPGSWNEQYODLCYEEGPFYKTFYVPK 60				
Qy	61 XLSQKHIRLYFAAVNTDCEVFLNGEKVGENHIEYLPEFVDVTGKVKSGENELRVVENRL 120				



```

Db      61 ELSQKHIRLYFAAVNTDCEVFLNGEKVGENHIEYLPFEVDVTGKVKSGENELRVVENRL 120
QY      121 KVGFPSPKVPDSGHTHTVGFSGFPANPDFFPYGGIIRPVLIEFTDHARILDIWDTSES 180
Db      121 KVGFPSPKVPDSGHTHTVGFSGFPANPDFFPYGGIIRPVLIEFTDHARILDIWDTSES 180
QY      181 EPEKKLGKVKKIEVSEAVGQEMTIKLGEEKKIRTSNRFVEGEFLENARFWSLEDPY 240
Db      181 EPEKKLGKVKKIEVSEAVGQEMTIKLGEEKKIRTSNRFVEGEFLENARFWSLEDPY 240
QY      241 LYLKVELKEDEYTLIDIGIRTIEMDEKRLYLNGKVPFLKGFGKHEEPVLGGTFYPLMI 300
Db      241 LYLKVELKEDEYTLIDIGIRTIEMDEKRLYLNGKVPFLKGFGKHEEPVLGGTFYPLMI 300
QY      301 KQPNLLKWINANSFRTSHYPYSEEWLADLRLGILVIDEAPHVGITRHYHNPETQKIAED 360
Db      301 KQPNLLKWINANSFRTSHYPYSEEWLADLRLGILVIDEAPHVGITRHYHNPETQKIAED 360
QY      361 NRRMIDRHKHNPVSIMSVANBPESNHPDAGEFFKALYETANEMDRTRPVVMSMDAP 420
Db      361 NRRMIDRHKHNPVSIMSVANBPESNHPDAGEFFKALYETANEMDRTRPVVMSMDAP 420
QY      421 DERTRDVALKYFDIVCNRYGYWYIYQRIEGLQALEKDIIEELYARHKRPIFTEFGAD 480
Db      421 DERTRDVALKYFDIVCNRYGYWYIYQRIEGLQALEKDIIEELYARHKRPIFTEFGAD 480
QY      481 AIAGIHYDPQMFSEYQAELEVEKTIKLLKKDYIIGTHVAFADFKTPQNVRRPILNHK 540
Db      481 AIAGIHYDPQMFSEYQAELEVEKTIKLLKKDYIIGTHVAFADFKTPQNVRRPILNHK 540
QY      541 GVFTDRQPKLVAVHLRLMSEV 563
Db      541 GVFTDRQPKLVAVHLRLMSEV 563

```

## RESULT 2

```

ID      Q97U11      PRELIMINARY;      PRT;      570 AA.
AC      Q97U11;
DT      01-OCT-2001 (Tremblrel. 18, Created)
DT      01-OCT-2001 (Tremblrel. 18, Last sequence update)
DE      Beta-glucuronidase (Gusb) (EC 3.2.1.31).
GN      Name=gusb; OrderedlocusNames=SSO3036;
OS      Sulfolobus solfataricus.
OC      Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OX      NCBI_TaxID=2287;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=ATCC 35092 / DSM 1617 / P2;
RX      MEDLINE=2132296; Pubmed=11427726;
RA      She Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,
RA      de Moore A., Erauso G., Fletcher C., Gordon P.M.K.,
RA      Heikamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,
RA      Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,
RA      Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,
RA      Garrett R.A., Regan M.A., Sensen C.W., Van der Oost J.,
RT      "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";
RL      Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).
DR      EMBL; AE006894; AAK43138.1; -.
DR      PIR; C90485; C90485.
DR      HSSP; P08236; 1BHG.
DR      GO; GO:0004566; F-beta-glucuronidase activity; IEA.
DR      GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR      InterPro; IPR008979; Gal_bind_like.
DR      InterPro; IPR006101; Glyco_hydro_2.
DR      InterPro; IPR006102; Glyco_hydro_2ig.
DR      InterPro; IPR006103; Glyco_hydro_2SB.
DR      InterPro; IPR006104; Glyco_hydro_2TM.
DR      Pfam; PF00703; Glyco_hydro_2; 1.
DR      Pfam; PF02836; Glyco_hydro_2_C; 1.

```

```

DR      Pfam; PF02837; Glyco_hydro_2_N; 1.
DR      PRINTS; PR00132; GLHYDRLASE2.
KW      Complete proteome; Glycosidase; Hydrolase.
SQ      SEQUENCE 570 AA; 66795 MW; DEB2FEC8050AF189 CRC64;

```

```

Query Match      33.7%; Score 1011; DB 2; Length 570;
Best Local Similarity 39.2%; Pred. No. 1.7e-58;
Matches 230; Conservative 92; Mismatches 191; Indels 74; Gaps 14;

```

```

QY      15 LNWVNIETSKDRP-----IAVGSWNEQYQDLCYEGEPFTYKTFYVPK 60
Db      11 LQGFWRKIDNENTGENGWYKLGSESDIYVPASWNEQNPKWDQSSGIAMYQKDLFVSN 70
QY      61 XLSQKHIRLYFAAVNTDCEVFLNGEKVGENHIEYLPFEVDVTGKVKSGENELRVVENRL 120
Db      71 DNGNRKAMVVEGAGYITKLWINGEYGTGHEGFTQFKPIKLV-----NEFNKIV---V 123
QY      121 KVGFPSPKVPDSGHTHTVGFSGFPAN-----FDPPYGGIIRPVLIEFTDHARILDIW 174
Db      124 KIDNTPSPY-----NLPPARDLNNAAFDFENYGGIHRPVYIEFVDECHVEDIT 171
QY      175 VDTSESEPEKKLGKVKKIEVSEAVGQEMTIKLGEEKKI-----RTSNRFVEGEFLEN 230
Db      172 VYT-----KSYGHILKEVI-LSECNQRFSLRFLKLVDEKGRVILNESSNEVEFKD--VNN 222
QY      231 ARFWSLEDPYLYPLKVELE-----KDEYTLIDIGIRTIEMDEKRLYLNGKVPFLKGFGKE 285
Db      223 VIPWSPDNPYLTYLIVEMTVYGVGNLKDVSVERIGFRDVEVKDKIYLNKGPILFGFGRHE 282
QY      286 EFPVLGGTFYPLMIKDFNLKWINANSFRTSHYPYSEEWLADLRLGILVIDEAP--HV 343
Db      283 DFPILGKFTYGAVLVVDPLYLMKIGANSFRTSHYPYSENEHLADDEMGLVILEBPLCY 342
QY      344 GITRHYHNPETQKI-----AEDNIRRMIDRHKHNPVSIMSVANBPESNHPDAEGF 394
Db      343 NISRYMSQEEIAKMGFDVKYFEKYRDTIKEMIRQHKNRPSVIMSVANBPESDIREVAEF 402
QY      395 FKALYETANEMDRTRPVVMSMDAPDERTRDVALKYFDIVCNRYGYWYIYQRIEGL 454
Db      403 IRREVELFKSLDSSRPVTFAS-----HRSVRDLALEYVDVISLNYHGWYTEMGDIDSGV 457
QY      455 QALEKDIEELVARRH-RKPIFVTEFGADAIAGIHYDPQMFSEYQAELEVEKTIKLLKKD 513
Db      458 KVAIAIELEEIHKKPPEKPIITTEFGADAIYGLHSDPPQMFSEYQSEMIRKYIEALREKD 517
QY      514 YITGTHVAFADFKTPQNVRRPILNHKGVFTDRQPKLVAVHLRLM 560
Db      518 YIVGFHIMNPAFRTQPNSRTIILNRKGIFTDRQPKLAAYVEELF 564

```

## RESULT 3

```

ID      Q8XP19      PRELIMINARY;      PRT;      599 AA.
AC      Q8XP19;
DT      01-MAR-2002 (Tremblrel. 20, Created)
DT      01-MAR-2002 (Tremblrel. 20, Last sequence update)
DE      Beta-glucuronidase.
GN      Name=bglR; OrderedlocusNames=CPE0147;
OS      Clostridium perfringens.
OC      Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OX      NCBI_TaxID=1502;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=13;
RX      MEDLINE=21664373; Pubmed=11792842;
RA      Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,
RA      Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.,
RT      "Complete genome sequence of Clostridium perfringens, an anaerobic
RT      flesh-eater.";
RL      Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).
DR      EMBL; AP003185; BAB79853.1; -.

```

DR HSSP; P08236; 1BHG.  
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.  
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.  
DR InterPro; IPR008979; Gal\_bind\_like.  
DR InterPro; IPR006101; Glyco\_hydro\_2.  
DR InterPro; IPR006102; Glyco\_hydro\_21g.  
DR InterPro; IPR006104; Glyco\_hydro\_25B.  
DR InterPro; IPR006103; Glyco\_hydro\_2TIM.  
DR Pfam; PF00703; Glyco\_hydro\_2; 1.  
DR Pfam; PF02836; Glyco\_hydro\_2\_C; 1.  
DR Pfam; PF02837; Glyco\_hydro\_2\_N; 1.  
DR PRINTS; PR00132; GLHYDRLASE2.  
DR PROSITE; PS00719; GLYCOSYL\_HYDROL\_F2\_1; 1.  
DR PROSITE; PS00608; GLYCOSYL\_HYDROL\_F2\_2; UNKNOWN\_1.  
KW Complete proteome.  
SQ SEQUENCE 599 AA; 68729 MW; 1AE4393869DAEF3B CRC64;

Query Match 32.6%; Score 977.5; DB 2; Length 599;  
 Best Local Similarity 36.9%; Pred. No. 3e-56;  
 Matches 227; Conservative 102; Mismatches 210; Indels 77; Gaps 16;

[illegible]

RESULT 4		
ID	Q8VNV4	PRELIMINARY; PRT; 599 AA.
AC	Q8VNV4;	
DT	01-MAR-2002	(TrEMBLrel. 20, Created)
DT	01-MAR-2002	(TrEMBLrel. 20, Last sequence update)
DT	01-OCT-2003	(TrEMBLrel. 25, Last annotation update)

DE Beta-glucuronidase.  
OS Clostridium perfringens.  
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
OC Clostridium.  
OX NCBI\_TaxID=1502;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=13;  
RX MEDLINE=21945361; PubMed=11948145;  
RA Briolat V., Reyssset G.;  
RT "Identification of Clostridium perfringens genes involved in the  
RT adaptive response to oxidative stress.";  
RL J. Bacteriol. 184:2333-2343(2002).  
EMBL; AJ420784; CAD12654.1; -.  
DR HSSP; P08236; IBHG.  
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.  
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.  
DR InterPro; IPR008979; Gal\_bind\_like.  
DR InterPro; IPR006101; Glyco\_hydro\_2.  
DR InterPro; IPR006102; Glyco\_hydro\_2ig.  
DR InterPro; IPR006104; Glyco\_hydro\_2SB.  
DR InterPro; IPR006103; Glyco\_hydro\_2TIM.  
DR Pfam; PF00703; Glyco\_hydro\_2; 1.  
DR Pfam; PF02836; Glyco\_hydro\_2\_C; 1.  
DR Pfam; PF02837; Glyco\_hydro\_2\_N; 1.  
DR PRINTS; PR00132; GLHYDRLASE2.  
DR PROSITE; PS00719; GLYCOSYL\_HYDROL\_F2\_1; 1.  
DR PROSITE; PS00608; GLYCOSYL\_HYDROL\_F2\_2; UNKNOWN\_1.  
DR SEQUENCE 599 AA; 68715 MW; BB538644AAED48 CRC64;

Query Match	32.4%;	Score 972.5;	DB 2;	length 599;
Best Local Similarity	36.7%;	Pred. No. 6.5e-56;		
Matches 226;	Conservative 102;	Mismatches 211;	Indels 77;	Gaps 16;

Oy		1	MWPQRNKKRFILINGVNMLEVTSKDR-----PIAVGSWNE--QYODLC	44
Dd		1	MLPYIITESRQLDLSGIWKFKLNGNGLTEELSKAPLEDTIEMAVSSYNDLVESQEVR	60
Oy		45	YEESPFYTKTFYVPKXLSQKHIRLYFAAVNTDCEVFLNGEKVGENHIEYLPEFVDVTGK	104
Dd		61	DHVGMWMYERNFTIPKTLINERIVLRFGSATHEAKYVLNGELLVEHKGFPTPEAEINDL	120
Oy		105	VKSGENELRVVENRLKVGGFPSPKVPDSGHTHTVGFSGSP-----PANFDFFPYCG	155
Dd		121	LVSQDNRLTVAVN-----IDETTLPGVLKEVEVDGKVIKNSVNFDFPNYAG	170
Oy		156	IIRPVLIETDHARILDIMWDTSSESEPEKKLGKVKKIEVSEEAQG-EMTIKLGEEEK	214
Dd		171	IHRPKIYTPPKSYIEDITIVTDKRENN---GYVNEY----QAVGRNIKVTIIDENN	223
Oy		215	IRTSNRFEVEGFLENARFWSLEDPYLPKLVELEKDEYTL-----IGRTISWDEKRL	269
Dd		224	IYAEGEKGEGKLTINNVMHEPMNAYLLKLVLELDDEEIIIDTYFEEFVRTVEVKDGKF	283
Oy		270	YLNCKPVLKFGFKHGEFPVLGGOTFYPLMIKD FNLLKMINANSFRSHYPYSEEWDLA	329
Dd		284	LINNPFYFKGFGKHEDSVYNGRGINEALINIKDFNLMKWIGANSFRISHYPYSEEIMRLA	343
Oy		330	DRUGLVIDEAPHVGITRYHYN-----PE-----TOKIAEDNIRMIDRHKN	371
Dd		344	DREGIVVIDETPAVGL--HLNFMA TGFGGDAPRKDWKEIGTKEAHERILRELVSBDKN	400
Oy		372	HBSVIMSVANEPBSNHPPDAEGFPKALYETANEMD-RTRPVVMVS--MMDAPDE-RTRDVA	428
Dd		401	HPCVVMMSVANEPDSDSEGAKEXHEBPLIKLTKELDPOKRPTVVITYLMSTPDRCCKVGDI V	460
Oy		429	LKYFDIVCVNRYGWYIYOGRIEGLQALEKDI EELYARHRK-PIVTEFGADAIAGIH Y	487
Dd		461	---DVLCLNRYGWYVAGGDLBEAKRMLEDELKGMEERC PKTPIMFTEYGADTVAGHD	516
Oy		488	DBPMFSESEYOABEVKTIRLLKKDYIIGHVAFADFKTPQNVARPILNHKGVFTRDR	547
Dd		517	TVPWFETEEOVEYYKANHEVMDCKNFGEQVWNPADFATSOGILIRVOGNKKGI FRER	576

QY 548 QPKLVAHLRLMSEV 563  
Db 577 KPKMIAHLRLMSEV 592

## RESULT 5

Q9AFA2 PRELIMINARY; PRT; 602 AA.  
AC Q9AFA2; 01-JUN-2001 (Tremblrel. 17, Created)  
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)  
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)  
DE Beta-glucuronidase.  
OS Staphylococcus sp. RLH1.  
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.  
OX NCBI\_TaxID=156489;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=RLH1;  
RA Jefferson R.A., Kilian A., Keese P.K.  
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=RLH1;  
RA Kilian A.  
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
DR HSSP; P08236; 1BHG.  
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.  
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.  
DR InterPro; IPR008979; Gal bind like.  
DR InterPro; IPR006101; Glyco\_hydro\_2.  
DR InterPro; IPR006102; Glyco\_hydro\_21g.  
DR InterPro; IPR006104; Glyco\_hydro\_2SB.  
DR InterPro; IPR006103; Glyco\_hydro\_2TIM.  
DR Pfam; PF00703; Glyco\_hydro\_2; 1.  
DR Pfam; PF02836; Glyco\_hydro\_2\_C; 1.  
DR Pfam; PF02837; Glyco\_hydro\_2\_N; 1.  
DR PRINTS; PR00132; GLHYDRLASE2.  
DR PROSITE; PS00719; GLYCOSYL\_HYDROL\_F2\_1; 1.  
DR PROSITE; PS00608; GLYCOSYL\_HYDROL\_F2\_2; 1.  
SQ SEQUENCE 602 AA; 68701 MW; 312AFCD11634D577 CRC64;

Query Match 31.0%; Score 929; DB 2; Length 602;  
Best Local Similarity 35.6%; Pred. No. 4.9e-53;  
Matches 221; Conservative 93; Mismatches 224; Indels 82; Gaps 15;

QY 1 MVRPQNRKRRPILILNGVNLV-----TSKDRPIAVPGSWNE--QYQDLIC 44  
Db 1 MLYPINTERTGVFDLNGVNLVFKLDYGLKEKWEYSKLTDTISMAVPSYNDIGVTKEIR 60  
QY 45 YEEGPTTYKTFYVPKLSQKHIRLYFAAVNTDCEVFLNGEKGVENHIEYLPEFVDVTGK 104  
Db 61 NHIGVWYEREFTVPAYLKDQRIVLRFGSATHKAIIVYNGELVVEHKGGLPFEEALINNS 120  
QY 105 VKSGENELRVVVENRLKVGFPSPKVPDSCGTHTVGFFGS-----FPPANDFFPY 154  
Db 121 LRDGMNRVAVDNL-----DSTLIPVGLYSEHHEGLGKVIKKNPFDFNYA 170  
QY 155 GIIRPVLIETDHAIRLIDVDTSESEPEKLGKVKKIEVSEAVGEMTIKLG--EE 212  
Db 171 GLHRPVKIYTPFTYVEDISVTDENGP---TGTVTYVDFQ---GKAETVKSVVDEE 223  
QY 213 KKIRTSNREVEGEFLEENARFWSLEDPIYPLKVELKDEYTLDI-----GIRTSMD 267  
Db 224 GKVAASTEGLSGNVEIPNVILMEPLNTYLYQIKVELVNDGLTIDVYEBPFGVTRVEVNDG 283  
QY 268 RLYLNGKPVFLKGFKEHEFPVLGGCTPYPLMIKDFNLKMINANSFRTSHYPYSEMLD 327  
Db 284 KFLINNKFPYFKGFGKHEDTPINGRGFEASNVMDFNILKMGANSFRTAHYPYSEELMR 343  
QY 328 LADRLGLIVIDEAPHVGTIRYHN-----PETOKIA-----EDNIRMIDR 368

Db 344 LADREGLVVIDETPAVG---HLNFMAITGLGGSERVSTWEKIRTFEHHQDVLRLVSR 400  
QY 369 HKNHPSVIMSVANEPESNHPDAEGFEKALYETANEND-RTRPVWVSMMDAPDERTDV 427  
Db 401 DKNHPSVIMSVIANEAATEEGAYEYFKPLVELTKELDPOKRPVTVILFVMATPE--TDK 458  
QY 428 ALKYFDIVCNRYGYMYIQGRIEGLQALEKDIIEELVARRH---KPIFTEFGADAIA 483  
Db 459 VAEILIDYALNRYNGWYFDGDLAARVHLR---QEFHAMNKRCPGKPMITEYGADTVA 515  
QY 484 GIHYDPQMFSEEQALVEKTIKRLILKKDYITGHWAFADFKTPQNVRRPILNHKGVF 543  
Db 516 GFHDIDPVMTFEEYQVYEQANHVFDEFENFVGQANFADFATSQGMRVQGNKGVF 575  
QY 544 TRDRQPKLVAHLRLMSEV 563  
Db 576 TRDRKPKLAHVRFRRTNTI 595

## RESULT 6

Q9AHJ8 PRELIMINARY; PRT; 598 AA.  
AC Q9AHJ8; 01-JUN-2001 (Tremblrel. 17, Created)  
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)  
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)  
DE Beta-glucuronidase.  
GN Name=gusA;  
OS Lactobacillus gasseri.  
OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;  
OX Lactobacillus.  
OX NCBI\_TaxID=1596;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ADH;  
RX MEDLINE=21141841; PubMed=11229918;  
RA Russell W.M., Klaenhammer T.R.;  
RT "Identification and cloning of gusA, encoding a new beta-glucuronidase  
from Lactobacillus gasseri ADH."  
RL Appl. Environ. Microbiol. 67:1253-1261 (2001).  
DR EMBL; AF305888; AAK07836.1; .  
DR HSSP; P08236; 1BHG.  
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.  
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.  
DR InterPro; IPR008979; Gal\_bind like.  
DR InterPro; IPR006102; Glyco\_hydro\_2.  
DR InterPro; IPR006101; Glyco\_hydro\_21g.  
DR InterPro; IPR006104; Glyco\_hydro\_2SB.  
DR InterPro; IPR006103; Glyco\_hydro\_2TIM.  
DR Pfam; PF00703; Glyco\_hydro\_2; 1.  
DR Pfam; PF02836; Glyco\_hydro\_2\_C; 1.  
DR Pfam; PF02837; Glyco\_hydro\_2\_N; 1.  
DR PRINTS; PR00132; GLHYDRLASE2.  
DR PROSITE; PS00719; GLYCOSYL\_HYDROL\_F2\_1; 1.  
SQ SEQUENCE 598 AA; 69762 MW; 5398F060B2DD887D CRC64;

Query Match 30.6%; Score 919.5; DB 2; Length 598;  
Best Local Similarity 34.1%; Pred. No. 2.1e-52;  
Matches 207; Conservative 110; Mismatches 227; Indels 63; Gaps 14;

QY 4 PQNRKRRPILILNGVNLVET-----SKDRP---IAPGSMNE--QYQDLCEE 47  
Db 7 PIONKRYFNTLMNGTWQFETDPSVGLDEGNKELPDPEMPVPGTFAELTTKDRKRYT 66  
QY 48 GPFTYKTFYVPKLSQKHIRLYFAAVNTDCEVFLNGEKGVENHIEYLPEFVDVTGKVK 107  
Db 67 GDFWYQKDFIPSLFKKELYIRFGSVTHRAKVFINGHEVGQHEGGLPFQVKISNYINY 126  
QY 108 GE-NELRVVVENRLKVGFP---SKVPDSCGTHTVGFFGSPPPANFDFFPYGGIIRPVLE 163  
Db 127 DQNRVTVLVNNELEKAIKPGCTEELDNGQ-----KLAQPYFDFFNYSGIMRNWLL 179

QY 164 FTDHARILDIWDTSESEPEKKLGKVKVIEVSEBAVGQ-EMTIKGEEKKIRTSNRFV 222  
Db 180 ALPOSQITNFKLN-----YOLANNKATITYNIEANNNAEFKVTLPFDNOKEVACATSKN 232  
QY 223 EGFEILENARFWSLEDPYLYPLKVELEK-----DEYTLDIGIRTSWDEKRLYLNGKPVF 277  
Db 233 TSSLTIKNPHLWSPNDPYSYKIKIEMLEDGKTVDYTDKIGIRTVXIVNDKILNNHPIY 292  
QY 278 LKGFGEKHEEPVLGGTFYPLMIKDENVLLKMINANSFRTSHYPYSEEWLADRLGILVI 337  
Db 293 LKGFGEKHEEDFNVLGKAVNESIIRKDYCEMKWIGANGFRSSHYPYAEWYQYADKYGFLII 352  
QY 338 DEAPHVGITRYHYN-----PETOKIAEDNIRRMIDRHKNHPSVIMWS 379  
Db 353 DEVPAVGLNRSITNELNVTNSNQSHPPASKTVPFLKCVHEQEIKEMIDRDQRHPSVIAMS 412  
QY 380 VANEPESNHPDAEGFFKALYETANEND-RTRPVVMVSMMDAPDERTRDVALKYPDIVCVN 438  
Db 413 LENEPESTTQESYDYFKDIFAFARXLDPQNRPYTGTLVMSGPK--VDKLHPLCDPVCCLN 470  
QY 439 RYGGWYIYOG-RIEBGLQALEKDIIEELY-ARHRKPIFVTEFGADAIAGIHYDPQMFSEB 496  
Db 471 RYGGWYVAGPEIVNAKKMLEDELQWQNLKLNKPEFVTEFGADTLSSSHRLPDEKWSQE 530  
QY 497 YQAEIVEXTIRLLKKDYIIGTHVAFADFKTPQNVRRPILNHKGVFTDRQPKLVAVL 556  
Db 531 YQNEYYQMYFDIFKXYPFICGELVWNPADFXTSEGIMRVGNDKGIFTDRDREPQIAFTL 590  
QY 557 RRLWSEV 563  
Db 591 KKRWQQL 597

RESULT 7  
Q6NL66 PRELIMINARY; PRT; 670 AA.  
AC Q6NL66;  
DT 05-JUL-2004 (TREMBLrel. 27, Created)  
DT 05-JUL-2004 (TREMBLrel. 27, Last sequence update)  
DT 05-JUL-2004 (TREMBLrel. 27, Last annotation update)  
DE RE15795P.  
GN Name=CG15117;  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxId=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Berkley;  
RA Stapleton M., Carlson J., Chavez C., Frise E., George R., Pacleb J.,  
RA Park S., Wan K., Yu C., Rubin G.M., Celniker S.;  
RL Submitted (APR-2004) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BT012475; AAS93746.1; -.  
DR InterPro; IPR008979; Gal\_bind\_1ke.  
DR InterPro; IPR006101; Glyco\_hydro\_2.  
DR InterPro; IPR006102; Glyco\_hydro\_2ig.  
DR InterPro; IPR006104; Glyco\_hydro\_2SB.  
DR InterPro; IPR006103; Glyco\_hydro\_2TIM.  
DR InterPro; IPR006025; Pept\_M\_zn\_BS.  
DR Pfam; PF00703; Glyco\_hydro\_2; 1.  
DR Pfam; PF02836; Glyco\_hydro\_2\_C; 1.  
DR Pfam; PF02837; Glyco\_hydro\_2\_N; 1.  
DR PRINTS; PR00132; GLHYDLASE2.  
DR PROSITE; PS00719; GLYCOSYL\_HYDROL\_F2\_1; 1.  
DR PROSITE; PS00608; GLYCOSYL\_HYDROL\_F2\_2; 1.  
DR PROSITE; PS00142; ZINC\_PROTEASE; UNKNOWN\_1.  
SQ SEQUENCE 670 AA; 77050 MW; 00AE0E67AE1D9E8C CRC64;

Query Match 30.2%; Score 906; DB 2; Length 670;  
Best Local Similarity 34.7%; Pred. No. 1.9e-51;  
Matches 212; Conservative 122; Mismatches 209; Indels 68; Gaps 17;

QY 1 MWBPQRNKKRIFILINGVWNL-----EVTSKDR---PIAVPGSWNEQ 39  
Db 48 MLYPRESETREVSLDGIWNFVRSDQANPTQGVREDEWYAKELSKSRPTIPMPVPAASYNDI 107  
QY 40 YQD-LCYEEGPFYKTTFFYVPKXLSQ-KHIRLYFAAVNTDCEVFLNGEKVGENHIEYLPF 97  
Db 108 TTDNLRDHVGTWYDRKFFVPRWSKQRIWLRFGSVHYEAAYWINGQKVVKHEMGLHPF 167  
QY 98 EVDVTGKVKSG-ENELRVVENRNLKVGFP---SKVPDSGTHTVGFFGSFPANFDFPP 152  
Db 168 BAEVTDLLSYGAEENRITVMCDNALIQTTVPQGRITVEPNDGWTI-----VQSYTFDFFN 222  
QY 153 YGGIIRPVLIETDARILDIWDTSESEPEKKLGKVKVIEVSEBAVGQ-----EMTIK 207  
Db 223 YAGIHRSVHLTYTPRTFIEEVEYTTNLSK-DATIGEVFSVSVGSAANEADNVLIQIAN 281  
QY 208 LGEEBKKI---RTSNRFVEGSEFILNARFW-----SLEDPLYPLKVELEK-----DEX 253  
Db 282 LYDKDGIIVANATSDQKLGKLOVNPVKPWPVYLMHSEBPGYLQLEIKLATNDELDDVY 341  
QY 254 TLDIGIRTSWDEKRLYLNGKPYFLKGFGEKHEEPVLGGTFYPLMIKDENVLLKMINANS 313  
Db 342 RLKVGIRTLSSWNGOQLINGKPYFRGFRHEDSDIRGKGLDNALMVRDFNLKWTIGANA 401  
QY 314 FRTSHYPYSEEWLADRLGILVIDEAPHVGITRYHYNPETOKIAEDNIRRMIDRHKNHP 373  
Db 402 YRTSHYPYSEESMGFADEHGIMTIDECPSVDE--NFSQELLGKHKSLSLEQLTHDRNHP 459  
QY 374 SVIMWSVANEPESNHPDAEGFFKALYETANENDRTRPVVMVSMMDAPDERTRDVALKYFD 433  
Db 460 SVVMWSIANEPRGVSADSYFELVANFTRSJDKTRPITAAIY-----SNTQDAGRSLD 515  
QY 434 IVCNRYGWIYQGRIEBGLQALEKDIIEELYA---RHRRKPIFVTEFGADAIAGIHYDPP 490  
Db 516 IISFNRYNAMYAGRLD---MITQNVIDEALAMNKRKYNKPIIMSEYGADTLEGLMQPA 572  
QY 491 QMFSEYQAEIVEXTIRL---LLKKDYIIGTHVAFADFKTPQNVRRPILNHKGVFTDR 547  
Db 573 YVWSEEFQTEVSRHFKAFADELRRKKGWPIGEPWNPADFXTAOSYTRVGNNKGVFTRAR 632  
QY 548 QPKLVAVLRR 558  
Db 633 QPKAAHLRR 643

RESULT 8  
AAS93746 PRELIMINARY; PRT; 670 AA.  
AC AAS93746;  
DT 23-APR-2004 (TREMBLrel. 27, Created)  
DT 23-APR-2004 (TREMBLrel. 27, Last sequence update)  
DT 23-APR-2004 (TREMBLrel. 27, Last annotation update)  
DE RE15795P.  
GN CG15117.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxId=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Berkley;  
RA Stapleton M., Carlson J., Chavez C., Frise E., George R., Pacleb J.,  
RA Park S., Wan K., Yu C., Rubin G.M., Celniker S.;  
RL Submitted (APR-2004) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BT012475; AAS93746.1; -.  
SQ SEQUENCE 670 AA; 77050 MW; 00AE0E67AE1D9E8C CRC64;

Query Match 30.2%; Score 906; DB 2; Length 670;  
Best Local Similarity 34.7%; Pred. No. 1.9e-51;  
Matches 212; Conservative 122; Mismatches 209; Indels 68; Gaps 17;  
QY 1 MVRPQRNKKRIFILINGVWNL-----EVTSKDR---PIAVPGSWNEQ 39

```

Db      48  MLYPRESETRVSLDGIWNVFRSDQANPTQGVNDEWYAKELSKSRPTIMPVPAASYNDI 107
Qy      40  YOD-LCYEGPFTYKTEFYPKLSQ-KHIRLYFAAVNTDCEVFLNGEKVGENHIEYLPF 97
Db      108  TTDNLRDHWGVWYDRKFFVPRSWKQDRIWLRFGSVHYEAYWINGQKVKGHMLPF 167
Qy      98  EVDVTGKVKSG-ENELRVVENRLKVGFP-----SKVPDSGTHTVGFGSFPANPDFFP 152
Db      168  EAEVTDLLSYGAENRITVMCDNALIQTTVPQGRITVEPNMGMTI-----VQSYTFDEFN 222
Qy      153  YGGIIRPVLIEFTDHARILDIWDTSESEPEKKLGKVKKIEVSEAVGQ-----EMTIK 207
Db      223  YAGIHRSVHLTYTTRTEIEEVEVTNLISK-DATIGEVFYSVSVGSAANADVLOIQAN 281
Qy      208  LGEEEEKI---RTSNRFEVEGFLEENARFW-----SLEDPLYPLKVELK-----DEY 253
Db      282  LYDKDGLVANATSDQKLGKLVNPKVPMWYLMHSEBPGYQLKIKLATNDELDDVY 341
Qy      254  TLDIGIRTISWDEKRLYLNGKPVFLKFGKHEEPVLGQGTFFPLMIKDFNLKWINANS 313
Db      342  RLKVGIRTLSNWSQOFLNGKPYFRGFRHEDSDIRGKGLNALMVRDFNLKWTIGANA 401
Qy      314  FRTSHYPYSEEWDLADRLGILVIDEAPHVGTIRVHYNPETOKIAEDNIRMRIDRKNHP 373
Db      402  YRTSHYPYSEESMQFADEHGMIDECPSVPTB--NFSQELGKHKSLQOLIHDRNHP 459
Qy      374  SVIMSVANEPESNHPDAEGFKALYETANEMDRTPRVVWVSMMDAPDERTDVALKYFD 433
Db      460  SVVWMSIANEPRTGSVSADSYELVANFTRSLDKTRPITAIAV-----SNTQDKAGRSLD 515
Qy      434  IVCVNRYYGYWYIQRIEGLQALEKDEIELYA---RHKKPIFVTEFGADAIAGIHYDP 490
Db      516  IISFNRYNAMYSNAGRLD--MITQNVIDEALAMNKRYPKIMSEYADTLEGLHMOPA 572
Qy      491  QMFSEEYQALVEKTIRL--LLKKDYIIGTHWAFADFKTPQNVRPILNHKGVFTRDR 547
Db      573  YWSEEFQTEVFSRHFKAFDELRRKGFIFGEFVWNFADFKTAQSYTRVGKKGVFTRAR 632
Qy      548  QPKLVAVHLRR 558
Db      633  QPKAAHLRR 643

```

RESULT 9

Q9V8R0 PRELIMINARY; PRT; 656 AA.

AC Q9V8R0; Q8T0G7;

DT 01-MAY-2000 (Tremblrel. 13, Created)

DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)

DT 05-JUL-2004 (Tremblrel. 27, Last annotation update)

DE CG15117-PA (LD04718P).

GN ORFNames=CG15117;

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydriidae; Drosophilidae; Drosophila.

OC NCB1\_TaxID=7227;

OX [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,

RA Brandon R.C., Rogers Y.H., Blazek R.G., Champe M., Pfeiffer B.D.,

RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,

RA Abril J.F., Agbayani A., An H.J., Andrews-Pfankoch C., Baldwin D.,

RA Ballem R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,

RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,

RA Borikova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,

RA Butris K.C., Butsam D.A., Butler H., Cadieu E., Center A., Chandra I.,

RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,

RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,

RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,

RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,

RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwan C.,

RA Jalali M., Kalush F., Kapten G.H., Ke Z., Kennison J.A., Ketchum K.A.,

RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

RA Liu X., Mattle B., McIntosh T.C., McLeod M.P., McPherson D.,

RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,

RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,

RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,

RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun B.,

RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,

RA Wang Z.Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,

RA Williams S.M., Woodagel, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,

RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,

RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,

RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,

RT "The genome sequence of Drosophila melanogaster.";

RL Science 287:2185-2195(2000).

RN [12]

RP SEQUENCE FROM N.A.

RX MEDLINE=22426065; PubMed=12537568;

RA Celniker S.E., Wheeler D.A., Krommiller B., Carlson J.W., Halpern A.,

RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,

RA George R.A., Hoskins R.A., Lavery T., Muzny D.M., Nelson C.R.,

RA Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,

RA Svirskas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,

RA Weinstein G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.,

RT "Finishing a whole-genome shotgun: release 3 of the Drosophila

RT melanogaster euchromatic genome sequence.";

RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).

RN [13]

RP SEQUENCE FROM N.A.

RX MEDLINE=22426070; PubMed=12537573;

RA Kaminker J.S., Bergman C.M., Krommiller B., Carlson J., Svirskas R.,

RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,

RA Ashburner M., Celniker S.E.,

RT "The transposable elements of the Drosophila melanogaster euchromatin:

RT a genomic perspective.";

RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).

RN [14]

RP SEQUENCE FROM N.A.

RX MEDLINE=22426069; PubMed=12537572;

RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,

RA Hradscky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,

RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,

RA Betencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.,

RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,

RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,

RA Lewis S.E.,

RT "Annotation of the Drosophila melanogaster euchromatic genome: a

RT systematic review.";

RL Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).

RN [15]

RP SEQUENCE FROM N.A.

RX MEDLINE=22426069; PubMed=12537572;

RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,

RA Hradscky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,

RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,

RA Betencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.,

RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,

RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,

RA Lewis S.E.,

RT "Annotation of the Drosophila melanogaster euchromatic genome: a

RT systematic review.";

RL Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).

RN [16]

RP SEQUENCE FROM N.A.

RX MEDLINE=22426069; PubMed=12537572;

RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,

RA Hradscky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,

RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,

RA Betencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.,

RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,

RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,

RA Lewis S.E.,

RT "Annotation of the Drosophila melanogaster euchromatic genome: a

RT systematic review.";

RL Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).

RN [17]

RP SEQUENCE FROM N.A.

RX MEDLINE=22426069; PubMed=12537572;

RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,

RA Hradscky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,

RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,

RA Betencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.,

RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,

RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,

RA Lewis S.E.,

RT "Annotation of the Drosophila melanogaster euchromatic genome: a

RT systematic review.";

RL Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).



RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,  
RA Nunoo J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,  
RA Yu C., Lewis S.E., Rubin G.M., Celniker S.,  
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AE003797; AAF57602.2; -.  
DR EMBL; AY069329; AAL39474.1; -.  
DR HSSP; P08236; 1BHG.  
DR FlyBase; FBgn0034417; CG15117.  
DR GO; GO:004553; F:hydrolyase activity, hydrolyzing O-glycosyl . . .; IEA.  
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.  
DR InterPro; IPR008979; Gal bind like.  
DR InterPro; IPR006101; Glyco\_hydro\_2.  
DR InterPro; IPR006102; Glyco\_hydro\_2ig.  
DR InterPro; IPR006104; Glyco\_hydro\_2SB.  
DR InterPro; IPR006103; Glyco\_hydro\_2TIM.  
DR InterPro; IPR006025; Pept\_M\_zn\_BS.  
DR Pfam; PF00703; Glyco\_hydro\_2; 1.  
DR Pfam; PF02836; Glyco\_hydro\_2\_C; 1.  
DR Pfam; PF02837; Glyco\_hydro\_2\_N; 1.  
DR PRINTS; PR00132; GLHYDRLASE2.  
DR PROSITE; PS00719; GLYCOSYL\_HYDROL\_F2\_1; 1.  
DR PROSITE; PS00608; GLYCOSYL\_HYDROL\_F2\_2; 1.  
DR PROSITE; PS00142; ZINC\_PROTEASE; UNKNOWN 1.  
SQ SEQUENCE 656 AA; 75166 MW; 34168D33C8F88A20 CRC64;

Query Match 30.2%; Score 905; DB 2; Length 656;  
Best local Similarity 34.7%; Pred. No. 2.1e-51;  
Matches 212; Conservative 122; Mismatches 209; Indels 68; Gaps 17;

QY 1 MVRPQNRKKRFFILNGVNL-----EVTSKDR--PIAVGSMNEQ 39  
DB 34 MLYPRESETRVRSLDGIWNPVRSQANPTQGVNDEWYAKELSKSRPTIPMPVPAASYNDI 93  
QY 40 YQD-LCYEEGPFTYKTTFYVPKLSQ-KHRLYFAVNTDCEVFLNGEKVGENHIEYLPF 97  
DB 94 TTDNLRDHVGIVWYDRKFFVPRWSKDQRIWLRFSGVHYEAYVWINGQKVKGHEMLPF 153  
QY 98 EVDVTGKVKSG-ENELRVVENRLKVGFP----SKVPDSGTHVGFSGFPFPAFDFPP 152  
DB 154 EAEVTDLLSYGAEENRITVMCDNALIQTTVPQGRITEVPNDGMTI----VQSYTFDEFN 208  
QY 153 YGGIIRPVLIEFTDHARILDIWDTSESEPEKKLGKVKKIEVSEAVGQ-----EMTIK 207  
DB 209 YAGIHRSVHLTYTPRTFIEEVEVTNLISK-DATVGEVFYSVSVGSAANEADVNLQIQAN 267  
QY 208 LGEEBEKKI---RTSNRFEVEGFLEENARFW----SLEDPLYLPLKVELK-----DEY 253  
DB 268 LYDKDGLVANATSDQKLGKGLQNPVKPMPWRYLMHSEPGYLQLEIKLATNDELDDVY 327  
QY 254 TLDIGIRTIWDEKRLYLNGKPYFLKGFGEHEFPVLGQGTFFYLMIKDFNLKWINANS 313  
DB 328 RLKVGIRTLSTNSQQLINGKRPYFRGFRHEDSDIRGKGLDNALMVRDFNLKWIGANA 387  
QY 314 FRTSHYPYSEEWLADRLGLIVIDEAPHVGTTRYHYNPETQKIAEDNIRRMIDRHKNHP 373  
DB 388 YRTSHYPYSEESMQFADEHGMIDECPSVDTE--NFSQELLGKHKSLLEQLHRDRNHP 445  
QY 374 SVIMMSVANEPESNHPDAEGFKALYETANEMDRTPVVMVMSMDAPDERTDVALKYFD 433  
DB 446 SVMMMSIANEPRTGSVSADSYFELVANFTRLSDKTRPITAIAV----SNTQDKAGRSLD 501  
QY 434 IVCNRRYYGWIYQGRIEEGLQALEKDIELVA---RHKKPIFVTEFGADAIAAGIHYDP 490  
DB 502 IISFNRYNAYWSNAGRDL--MITQNVIDEAIAWKNRYNKPIMINSEYGADTLEGLMQPA 558  
QY 491 QMFSEEYQAELEVTIRL---LKKDYIIGTHVAFADFKTPQNVRRPILNHKGVFTDR 547  
DB 559 YWSEEFQTEVFSRHFKAFFDELKRGKGFIGEFVNNFADFKTAQSYTRVGKNGKGVFTRR 618  
QY 548 QPKLVAHLRR 558  
DB 619 QPKAAHLRLRK 629

RESULT 10  
Q8MB7 PRELIMINARY; PRT; 670 AA.  
AC Q8MB7;  
DT 01-OCT-2002 (Tremblrel. 22, Created)  
DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)  
DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)  
DE CG15117-PB.  
GN ORFNames=CG15117;  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxId=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20196006; Pubmed=10731132;  
RA Adams M.D., Celniker S.B., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.B., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.H., Blazer R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,  
RA Abril J.F., Agbayani A., An H.J., Andrews-Pfankoch C., Baldwin D.,  
RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,  
RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Fosler C., Gabriellian A.E., Gary N.S., Gelbart W.M., Glasser K.,  
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwan C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Laeko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Pui V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.Y., Wassarman D.A., Weinstein G.M., Weissbach J.,  
RA Williams S.M., Woodgett, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,  
RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers B.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of Drosophila melanogaster.";  
RL Science 287:2185-2195 (2000).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=22426065; Pubmed=12537568;  
RA Celniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,  
RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,  
RA George R.A., Hoskins R.A., Laverly T., Muzny D.M., Nelson C.R.,  
RA Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,  
RA Svirskaas R., Taber P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,  
RA Weinstein G., Scherer S.E., Myers B.W., Gibbs R.A., Rubin G.M.;  
RT "Finishing a whole-genome shotgun: release 3 of the Drosophila  
melanogaster euchromatic genome sequence.";  
RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).  
RN [3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=22426070; Pubmed=12537573;  
RA Kaminker J.S., Bergman C.M., Kronmiller B., Carlson J., Svirskaas R.,  
RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,

RA Ashburner M., Celniker S.E.;  
RT "The transposable elements of the Drosophila melanogaster euchromatin:  
RT a genomics perspective."; Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
RL  
RN [4]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=22426069; PubMed=12537572;  
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,  
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,  
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,  
RA Belencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.,  
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,  
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,  
RA Lewis S.E.;  
RT "Annotation of the Drosophila melanogaster euchromatic genome: a  
RT systematic review."; Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).  
RL  
RN [5]  
RP SEQUENCE FROM N.A.  
RG FLYBASE;  
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
RN [6]  
RP SEQUENCE FROM N.A.  
RG FLYBASE;  
RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AE003797; AAM68434.1; -.  
DR HSSP; P08236; 1BHG.  
DR FlyBase; FBgn0034417; CG15117.  
DR GO; GO:0004553; F:hydrolyase activity, hydrolyzing O-glycosyl . . .; IEA.  
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.  
DR InterPro; IPR008979; Gal\_bind\_like.  
DR InterPro; IPR006101; Glyco\_hydro\_2.  
DR InterPro; IPR006102; Glyco\_hydro\_2ig.  
DR InterPro; IPR006104; Glyco\_hydro\_2SB.  
DR InterPro; IPR006103; Glyco\_hydro\_2TIM.  
DR InterPro; IPR006025; Pept\_M\_Zn\_BS.  
DR Pfam; PF00703; Glyco\_hydro\_2; 1.  
DR Pfam; PF02836; Glyco\_hydro\_2\_C; 1.  
DR Pfam; PF02837; Glyco\_hydro\_2\_N; 1.  
DR PRINTS; PR00132; GLHYDRLASE2.  
DR PROSITE; PS00719; GLYCOSYL\_HYDROL\_F2\_1; 1.  
DR PROSITE; PS00608; GLYCOSYL\_HYDROL\_F2\_2; 1.  
DR PROSITE; PS00142; ZINC\_PROTEASE; UNKNOWN 1.  
SQ SEQUENCE 670 AA; 77036 MW; 7A9F23351E458C90 CRC64;

	Query Match	30.2%;	Score 905;	DB 2;	Length 670;	
	Best Local Similarity	34.7%;	Pred. No. 2.2e-51;			
	Matches 212;	Conservative 122;	Mismatches 209;	Indels 68;	Gaps 17;	
QY	1 MVRPQRNKKRFILLINGVWNL-----EVTSKDR--PIAVPGSWNEQ	39				
Db	48 MLYPRESETRVRSLDGIWNFVRSDQANPTQGVDEWYAKELSKSRPTIPMPVPASYNDI	107				
QY	40 YQD-LCYEEGFTTKTFYVPPKLSQ-KHIRLYFAAVNTDCEVFLNGEKVGENHIEYLPE	97				
Db	108 TTDNLRDHVGTWYDRKFVFPRSMKQRIWLRFSGSVHYEAAYIINGQKVVKHEMGLPLP	167				
QY	98 EVDVTGKVKKS-G-ENELRVVENRLKVGFP----SKVPDSGTHTVGFSGSPRANPDFFP	152				
Db	168 EAETDLTLYGAENRITVMCDNALIQTTVPQGRITEVPNDGWTI-----VGSYTFDFFN	222				
QY	153 YGGIRPVLIIEFTDHARILIDIWDVTSESEPEKKLQKVKVKEIVSEAVGQ-----EMTIK	207				
Db	223 YAGIHRSVHLTYTPRTPFIEEVEVTNLISK-DATVGEVFFYSVSGSAANEADNVLIQIAN	281				
QY	208 LGEEKKI---RTSNRFVEGEFILENARFW-----SLEDPYLYPLKVELER-----DEY	253				
Db	282 LYDKDGILVANATSDDKLGGKLGQNVPKPMWPYLMHSEPGYALQLIKLATNDELIDVY	341				
QY	254 TLDIGIRTISSWDEKRLYLNGRPVFLKGFGKHIEEPVLQGQTFEYPLMIKDENVLLKWINANS	313				
Db	342 RLKVGIRTLSSWSQQFLINKGPVYFRFGFGRHEDSDIRGKGLDNALMVREDENLLKWIGANA	401				

[illegible]

RESULT	ID	BGLR_MOUSE	STANDARD	PRT	648 AA.
AC	P12265	Q61601	Q64473	Q64474	
DT	01-OCT-1989	Rel. 12	Created		
DT	01-OCT-1989	Rel. 12	Last sequence update		
DT	05-JUL-2004	Rel. 44	Last annotation update		
DE	Beta-glucuronidase precursor (EC 3.2.1.31).				
GN	Name=Gusb; Synonyms=Gus, Gus-s;				
OS	Mus musculus (Mouse).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
OX	NCBI_TaxID=10090;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=88085188; PubMed=2891607;				
RA	Gallagher P.M., D'Amore M.A., Lund S.D., Elliott R.W., Pazik J.,				
RA	Hohman C., Korfhagen T.R., Ganschow R.E.;				
RT	"DNA sequence variation within the beta-glucuronidase gene complex				
RT	among inbred strains of mice.";				
RL	Genomics 1:145-152(1987).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=88284700; PubMed=3397060;				
RA	Gallagher P.M., D'Amore M.A., Lund S.D., Ganschow R.E.;				
RT	"The complete nucleotide sequence of murine beta-glucuronidase mRNA				
RT	and its deduced polypeptide.";				
RL	Genomics 2:215-219(1988).				
RN	[3]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=89062453; PubMed=3196706;				
RA	D'Amore M.A., Gallagher P.M., Korfhagen T.R., Ganschow R.E.;				
RT	"Complete sequence and organization of the murine beta-glucuronidase				
RT	gene.";				
RL	Biochemistry 27:7131-7140(1988).				
RN	[4]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=YBR, and C3H/HeJ, TISSUE=Sperm;				
RX	MEDLINE=89384641; PubMed=2779578;				
RA	Wawrzyniak C.J., Gallagher P.M., D'Amore M.A., Carter J.E., Lund S.D.,				
RA	Rinchik E.M., Ganschow R.E.;				
RT	"DNA determinants of structural and regulatory variation within the				
RT	murine beta-glucuronidase gene complex.";				
RL	Mol. Cell. Biol. 9:4074-4078(1989).				
RN	[5]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=88216590; PubMed=2835664;				
RA	Funkhstein B., Leary S.L., Stein J.C., Catterall J.F.;				
RT	"Genomic organization and sequence of the Gus-s alpha allele of the				
RT	murine beta-glucuronidase gene.";				
RL	Mol. Cell. Biol. 8:1160-1168(1988).				
CC	-1- FUNCTION: Plays an important role in the degradation of dermatan				

```
CC and keratan sulfates.
CC -1- CATALYTIC ACTIVITY: A beta-D-glucuronoside + H(2)O = an alcohol +
CC D-glucuronate.
CC -1- SUBUNIT: Homotetramer.
CC -1- SUBCELLULAR LOCATION: Lysosomal.
CC -1- SIMILARITY: Belongs to family 2 of glycosyl hydrolases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; J03047; AAA37696.1; -
DR EMBL; J02836; AAA98623.1; -
DR EMBL; M63836; AAA63309.1; -
DR EMBL; M28540; AAA63307.1; -
DR EMBL; M28541; AAA63308.1; -
DR EMBL; M19279; AAA37697.1; -
DR PIR; A32576; A32576.
DR HSSP; P08236; 1BHG.
DR MGI; MGI:95872; Gue.
DR InterPro; IPR008979; Gal_bind_1like.
DR InterPro; IPR006101; Glyco_hydro_2.
DR InterPro; IPR006102; Glyco_hydro_219.
DR InterPro; IPR006104; Glyco_hydro_258.
DR InterPro; IPR006103; Glyco_hydro_2TIM.
DR Pfam; PF00703; Glyco_hydro_2; 1.
DR Pfam; PF02836; Glyco_hydro_2_C; 1.
DR Pfam; PF02837; Glyco_hydro_2_N; 1.
DR PRINTS; PR00132; GLHYDRLASE2.
DR PROSITE; PS00719; GLYCOSYL_HYDROL_F2_1; 1.
DR PROSITE; PS00608; GLYCOSYL_HYDROL_F2_2; 1.
KM Glycoprotein; Glycosidase; Hydrolase; Lysosome; Signal.
FT SIGNAL 1 22
FT CHAIN 23 648 Beta-glucuronidase.
FT ACT SITE 447 447 Proton donor (By similarity).
FT CARBOHYD 172 172 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 416 416 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 591 591 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 627 627 N-linked (GlcNAc...) (Potential).
FT VARIANT 87 87 T -> I (in strain C3H/HeJ).
FT VARIANT 233 233 I -> T (in allele GUS-SA).
FT VARIANT 265 265 D -> G (in strain YBR and strain C3H/HeJ).
FT VARIANT 320 320 V -> I (in strain YBR and strain C3H/HeJ).
FT VARIANT 428 428 E -> K (in allele GUS-SA).
FT VARIANT 616 616 F -> L (in allele GUS-SA).
SQ SEQUENCE 648 AA; 74239 MW; 3D8C65A5DB3B96D6 CRC64;

Query Match 30.1%; Score 902.5; DB 1; Length 648;
Best Local Similarity 35.7%; Pred. No. 3.1e-51;
Matches 223; Conservative 103; Mismatches 207; Indels 91; Gaps 21;

QY 1 MVRPORNKKRIFILINGWNLV-TSKDR-----PIAVPGSWNEQ 39
DB 27 MLEPKESPRLKALDGLMFRADLSNNRLQGEQWYRQPLRESGPLDMVPSSFNDI 86
QY 40 YQDLCYEE--GPFTYKTFVVPKLSQ--KHRLYFAAVNTDCEVLNGEKVGENHIEY 94
DB 87 TQEAALRDFIGWWYEREARLBRRTQDTDMRVLRINSAHYAVVWNGIHVEHEGCH 146
QY 95 LPFEVDVTGKYVSG---ENELRVVENRLKYGGFPSPKVP-----DSGTHVGFSGSF 143
DB 147 LPFEADISKLVQSGPLTTCRTIAINNTLT-----PHTLPPGTIVYKTDTSWPKGYF--V 200
QY 144 PPANDFPFPYGGIIRPVLIEFTDHARILDIVDTSESEPEKKLGKVK--VKIEVSEAVG 201
DB 201 QDTSFDFFNAGLHRSVLVLTPTTYIDITVTN---VEQDIGLVTYWISVGSEHF-- 255
```

```
QY 202 QEMTIKLGEEBKIRTSNRFVEGEFLENARFW----SLEDP-YLYPLKV-----ELEKD 251
DB 256 -QLEVQLDEDEGKVVAHGTCNQQLQVPSANLWVPLMHEHPAYMSLEVKYTTTESVTD 314
QY 252 EYTLDIGIRTSWDEKRLYLNGKPVFLKGFGEHSEFPVLGGTFYPLMIKDFNLKWINA 311
DB 315 YTLPVGIRIVAVTKSKFLNGKPFYFQGVNKHEDSDIRGFDWPLLVKDFNLLRWLGA 374
QY 312 NSFRTSHYPYSEEWLADRLGILVIDEAPHVIGT-----RYHNPETOKIAEDN 361
DB 375 NSFRTSHYPYSEEWLQCDRYGIVVIDECPGVILPQSGFNGESLRHNL-----EVMEEL 429
QY 362 IRRMIDRHKNPSVIMWSVANEPPESNHPDAGEFFKALYETANEMDRTPVWVS--MMDA 419
DB 430 VR---RDKNHPAVVWMSVANEPSSALKPAAVYFKTLITHTKALDLTRPVFVSNKAYDA 485
QY 420 PDERTRDVALKYFDIVCVNRYGYIYGRIEGLQALEKDI EELYARHRKPIFVTEFGA 479
DB 486 -----DLGAPYVDVICVNSYFSWYHDYGHLEVIQPOLNSQFENWYKTHQKPIIQSEYGA 539
QY 480 DALAGIHYDPQMFSEEQAEIVKTI RL--LKDYIIGTHVWAFADFKTPQNVRRPIL 537
DB 540 DAIPGIHEDPPRMFSEEQAVLENYHSVLDQKRKEYVVGELIWNFADFMTNQSPLRVIG 599
QY 538 NHKGVFTDRQPKLVAVHLR-RLW 560
DB 600 NKKGIFFTRQRPKTSAFILERYW 623

RESULT 12
ID Q6IR10 PRELIMINARY; PRT; 648 AA.
AC Q6IR10;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Beta-glucuronidase.
GN Name=Gus;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NMRI; TISSUE=Mammary tumor. WAP-Tag model. 5 months old;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.U., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahney J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywicki M.I., Skalska U., Smallus D.E., Scherch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=NMRI; TISSUE=Mammary tumor. WAP-Tag model. 5 months old;
RA Strausberg R.;
RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC071226; AA071226.1; -
DR InterPro; IPR008979; Gal_bind_1like.
```

DR InterPro; IPR006101; Glyco\_hydro\_2.  
DR InterPro; IPR006102; Glyco\_hydro\_219.  
DR InterPro; IPR006104; Glyco\_hydro\_25B.  
DR InterPro; IPR006103; Glyco\_hydro\_2TIM.  
DR Pfam; PF00703; Glyco\_hydro\_2; 1.  
DR Pfam; PF02836; Glyco\_hydro\_2\_C; 1.  
DR Pfam; PF02837; Glyco\_hydro\_2\_N; 1.  
DR PRINTS; PR00132; GLHYDRLASE2.  
DR PROSITE; PS00719; GLYCOSYL\_HYDROL\_F2\_1; 1.  
DR PROSITE; PS00608; GLYCOSYL\_HYDROL\_F2\_2; 1.  
SQ SEQUENCE 648 AA; 74195 MW; CCD8F84C3CD6C498 CRC64;

Query Match 30.0%; Score 899.5; DB 2; Length 648;  
Best Local Similarity 35.9%; Pred. No. 4.9e-51;  
Matches 224; Conservative 101; Mismatches 208; Indels 91; Gaps 21;

QY 1 MVRPQNRKRFILINGVNLV-TSKDR-----PIAVPGSWNEQ 39  
DB 27 MLFPKESPSRELKALDGLMFRADLSNNRLQGFQGWYRQPLRESGPVLDMPVSSPNDI 86  
QY 40 YQDLCEEE--GPFTYKTFPVVPKLSQ--KHRLYFAAVNTDCEVFLNGEKGVENHIEY 94  
DB 87 TQEAALRDFIGWVYEREAIIIPRWOTDTMRVLRINSAHYAVVWNGIHVEHEGGH 146  
QY 95 LPFEVDVTGKVGSG--ENELRVVENRLKVGSPSKVP-----DSGTHVGFSGSF 143  
DB 147 LPFEADISKLVQSGPLTTCRTITAINNTLT---PHTLPPTIVYKTDTSMPKGYF--V 200  
QY 144 PRANDFPFYGGIIRPVLIETDHAILEDVDTSESEPEKLGKVK--VKIEVSEAVG 201  
DB 201 QDTSFDFPNYAGLHRSVLTPTTYIDITVITN--VEQDIGLVYVWISVQSEHF-- 255  
QY 202 QEMTIKLGEEKKIRTSNFEVEGEFILENARFV---SLBDP-VLYPLKY----ELEKD 251  
DB 256 -QLEVQLDEGGKVVAHGNGOQLQVPSANLWVPLMHEHPAYMYSLEVAVTTTESVTD 314  
QY 252 EYTLDIGRTISWDEKRLYNGKPVFLKFGKHEEPVVGQGTFFPLMKDNLKWTNA 311  
DB 315 YTLPIGIRTVAVTSSKFLINGKPFYFGQVNHEDSDIRGKGFDWPLVXDFNLRLWLG 374  
QY 312 NSFRTSHYPISEEWLADRLGILVIDEAPHVGIT-----RYHNPETOKIAEDN 361  
DB 375 NSFRTSHYPISEEVLOLCRYGIVVIDECPGVGIVLPQSGFNESLRHHI----EVMEE 429  
QY 362 IRRMIDRKHNPVIMSVANEPESNHPAEGFFKALYETANEMDRTRPVVMS--MMDA 419  
DB 430 VR---RDKNHPAVVMSVANEPSALKAAYFFKTLITHTKALDLTRPVTFVSNAXYDA 485  
QY 420 PDERTDVALKYFDIVCVNRYYGYTYQGRIEBGLQALEKDIELVYARHKPIFVTEGA 479  
DB 486 -----DLGAPYVDVICVNSYFSWYHDYHLEVIQPOLNSOFENWYKTHQKPIIQSEYGA 539  
QY 480 DAIAIGHYDPQPMFSESEYQALVEKTRIL--LKKDYITIGTHWAFADFKTPQNVRRPIL 537  
DB 540 DAIPGIHEDPPRMFSESEYQAVLENHNSVLDQKRKEYVVGELIMNFAFMNTNQSPRLVIG 599  
QY 538 NHKGVFTDRQPKLVAVLR-RLW 560  
DB 600 NKKGIFTRQRPKTSAPILRERYW 623

RESULT 13  
BGLR\_FELCA STANDARD; PRT; 651 AA.  
AC 097524;  
DT 15-JUL-1999 (Rel. 38, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DT 05-JUL-2004 (Rel. 44, Last annotation update)  
DE Beta-glucuronidase precursor (EC 3.2.1.31).  
GN Name=GUSB;  
OS Felis silvestris catus (Cat).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.

OX NCBI\_TaxID=9685;  
RN [1]  
RP SEQUENCE FROM N.A., AND VARIANT MPS VII LYS-351.  
RC TISSUE=Liver;  
RX MEDLINE=99296826; Pubmed=10366443; DOI=10.1006/geno.1999.5825;  
RA Fyfe J.C., Kurzhals R.L., Lassaline M.E., Henthorn P.S., Alur P.R.,  
RA Wang P., Wolfe J.H., Giger U., Haskins M.E., Patterson D.F., Sun H.,  
RA Jain S., Yuhki N.;  
RT "Molecular basis of feline beta-glucuronidase deficiency: an animal  
RT model of mucopolysaccharidosis VII.";  
RL Genomics 58:121-128 (1999).  
CC -!- FUNCTION: Plays an important role in the degradation of dermatan  
CC and keratan sulfates.  
CC -!- CATALYTIC ACTIVITY: A beta-D-glucuronoside + H(2)O = an alcohol +  
CC D-glucuronate.  
CC -!- SUBUNIT: Homotetramer (By similarity).  
CC -!- SUBCELLULAR LOCATION: Lysosomal.  
CC -!- DISEASE: Defects in GUSB are the cause of mucopolysaccharidosis  
CC type VII (MPS VII), an inherited disease reported in humans, mice,  
CC cats, and dogs.  
CC -!- SIMILARITY: Belongs to family 2 of glycosyl hydrolases.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AF012423; AAD01498.1; -.  
DR EMBL; AF012424; AAD01499.1; -.  
DR HSSP; P08236; 1BHG.  
DR InterPro; IPR008979; Gal\_bind\_like.  
DR InterPro; IPR006101; Glyco\_hydro\_2.  
DR InterPro; IPR006102; Glyco\_hydro\_219.  
DR InterPro; IPR006104; Glyco\_hydro\_25B.  
DR InterPro; IPR006103; Glyco\_hydro\_2TIM.  
DR Pfam; PF00703; Glyco\_hydro\_2; 1.  
DR Pfam; PF02836; Glyco\_hydro\_2\_C; 1.  
DR Pfam; PF02837; Glyco\_hydro\_2\_N; 1.  
DR PRINTS; PR00132; GLHYDRLASE2.  
DR PROSITE; PS00719; GLYCOSYL\_HYDROL\_F2\_1; 1.  
DR PROSITE; PS00608; GLYCOSYL\_HYDROL\_F2\_2; 1.  
KW Disease mutation; Glycoprotein; Glycosidase; Hydrolase; Lysosome;  
KW signal.  
FT SIGNAL 1 22 By similarity.  
FT CHAIN 23 651 Beta-glucuronidase.  
FT ACT\_SITE 450 450 Proton donor (By similarity).  
FT CARBOHYD 172 172 N-linked (GlcNAc...) (Potential).  
FT CARBOHYD 419 419 N-linked (GlcNAc...) (Potential).  
FT CARBOHYD 630 630 N-linked (GlcNAc...) (Potential).  
FT VARIANT 351 351 E -> K (in MPS VII).  
SQ SEQUENCE 651 AA; 74609 MW; 2AE30884B70D4232 CRC64;

Query Match 29.9%; Score 898; DB 1; Length 651;  
Best Local Similarity 36.4%; Pred. No. 6.2e-51;  
Matches 224; Conservative 98; Mismatches 212; Indels 82; Gaps 21;

QY 4 PQRNKKRFILINGVNLV-TSKDR-----PIAVPGSWNEQYOD 42  
DB 34 PSRERKE----LNGLMSFRADPSNRQGFQGWYRTPPLRESGPTLMDMPVSSPNDVGDQ 89  
QY 43 --LCYEEGPFYKTFPVVPKLSQ--KHRLYFAAVNTDCEVFLNGEKGVENHIEYLPF 97  
DB 90 RQLRSFVGWVYEREAITLPQRTQDGLTRVLRIGSAHYAVVWNGVHAHEGGHLPF 149  
QY 98 EVDVTGKVGSG--ENELRVVENRLKVGSPSKVP-----DSGTHVGFSGSFPPA 146  
DB 150 EADISKLVQSGPLASCRITIAINNTLT---PHTLPPTIVYQDTSKYPKGYF--VQNI 203  
QY 147 NPDFFPYGGIIRPVLIETDHAILEDVDTSESEPEKLGKVKVKIEVSEAVGQEMTI 206

DB 204 NDEFFNYAGLRHRLVLLTPTTYIDITISTSVNQ-DTGLVDYQIFVEGGEHF---QLEV 259  
QY 207 KLGEEBKIRTSNRFEVEGEFILENARFW---SLEDP-YLYPLKVELE-----KDEX 253  
DB 260 RLDEEGKVAQGTGGRGQLQVPAHLWMPYLMHEHPAYLYSLEVRLTAQTAAGSVSDFY 319  
QY 254 TLDIGIRTIISWDEKRLYLNGKPVFLKFGKHEEPVLQGGTFYPLMIKDFNLKMINANS 313  
DB 320 TLPVGIRTVAVTEHQFLINGKPFYHGVNKHEDADIRGKGFDWPLLVKDFNLLRWLGANA 379  
QY 314 FRTSHYPYSEEWLADRLGLVVIDEAPHVGI--TRYHNPETO---KIAEDNIRRMIDR 368  
DB 380 FRTSHYPYAEVWQLCDRYGIVVIDESPGVGLVYESYSNVSLQHLEVMEEIVR---R 435  
QY 369 HKNHPSVIMMSVANEPESNHPDAEGFFKALYETANEMDRTRPVVMVSMMDAPDERTRDVA 428  
DB 436 DKNHPAVVMMSVANEPAFLKRPAGYVFKTLIAHTKALDPSRPVTFVT---NSNYEADLG 491  
QY 429 LKXFEDIVCNRYGYWYIYQGRIEGLQALEKDIIEELYARHKRPIFVTEFGADAIAGIHYD 488  
DB 492 APYVDVICVNSYSWYHDYGHMEYIQLQLATQFENWYRTYQKPIIQSEYGADRTIAGFHQD 551  
QY 489 PQMFSEBYQALVEKTIKRL---KKDYIIGTHVMAFADFPTQNVRRPILNHKGVFTR 545  
DB 552 PMLFSEBYQKGLLEQ-YHLVLDQKRKEXVVGELIMNFADFMFTNQSPQRVMGNKKGIIFTR 610  
QY 546 DRQPKLVAAHVLR-RLW 560  
DB 611 QRPKGAAPLRLERYW 626

RESULT 14  
BGLR\_RAT  
ID\_BGLR\_RAT STANDARD; PRT; 648 AA.  
AC P06760;  
DT 01-JAN-1988 (Rel. 06, Created)  
DT 01-JAN-1988 (Rel. 06, Last sequence update)  
DT 05-JUL-2004 (Rel. 44, Last annotation update)  
DE Beta-glucuronidase precursor (EC 3.2.1.31).  
GN Name=Gusb; Synonyms=Gus;  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Preputial gland;  
RX MEDLINE=87016933; PubMed=3463967;  
RA Nishimura Y., Rosenfeld M.G., Kreibich G., Gubler U., Sabatini D.D.,  
RA Adesnik M., Andy R.;  
RT "Nucleotide sequence of rat preputial gland beta-glucuronidase cDNA  
RT and in vitro insertion of its encoded polypeptide into microsomal  
RT membranes";  
RL Proc. Natl. Acad. Sci. U.S.A. 83:7292-7296(1986).  
RN [2]  
RP SEQUENCE OF 14-648 FROM N.A.  
RC TISSUE=Liver;  
RX MEDLINE=88183378; PubMed=335537;  
RA Powell P.P., Kyle J.W., Miller R.D., Pantano J., Grubb J.H., Sly W.S.;  
RT "Rat liver beta-glucuronidase. cDNA cloning, sequence comparisons and  
RT expression of a chimeric protein in COS cells.";  
RL Biochem. J. 250:547-555(1988).  
CC -1- FUNCTION: Plays an important role in the degradation of dermatan  
CC and keratan sulfates.  
CC -1- CATALYTIC ACTIVITY: A beta-D-glucuronoside + H(2)O = an alcohol +  
CC D-glucuronate.  
CC -1- SUBUNIT: Homotetramer.  
CC -1- SUBCELLULAR LOCATION: Lysosomal.  
CC -1- PTM: Undergoes a post-transcriptional proteolytic cleavage near  
CC its C-terminal end, which reduces its size by approximately 3 kDa.  
CC The site of this cleavage has as yet not been determined.  
CC -1- SIMILARITY: Belongs to family 2 of glycosyl hydrolases.  
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M13962; AAA41228.1; -.  
DR EMBL; Y00717; CAA68705.1; -.  
DR PIR; A25047; A25047.  
DR HSSP; P08236; 1BHG.  
DR RGD; 2772; Gusb.  
DR InterPro; IPR008979; Gal\_bind\_like.  
DR InterPro; IPR006101; Glyco\_hydro\_2.  
DR InterPro; IPR006102; Glyco\_hydro\_21g.  
DR InterPro; IPR006104; Glyco\_hydro\_2SB.  
DR InterPro; IPR006103; Glyco\_hydro\_2TIM.  
DR Pfam; PF00703; Glyco\_hydro\_2; 1.  
DR Pfam; PF02836; Glyco\_hydro\_2\_C; 1.  
DR Pfam; PF02837; Glyco\_hydro\_2\_N; 1.  
DR PRINTS; PR00132; GLHYDLASE2.  
DR PROSITE; PS00719; GLYCOSYL\_HYDROL\_F2\_1; 1.  
DR PROSITE; PS00608; GLYCOSYL\_HYDROL\_F2\_2; 1.  
KW Glycoprotein; Glycosidase; Hydrolase; Lysosome; Signal.  
FT SIGNAL 1 22  
FT CHAIN 23 648 Beta-glucuronidase.  
FT ACT\_SITE 447 447 Proton donor (By similarity).  
FT CARBOHYD 172 172 N-linked (GlcNAc...) (Potential).  
FT CARBOHYD 416 416 N-linked (GlcNAc...) (Potential).  
FT CARBOHYD 591 591 N-linked (GlcNAc...) (Potential).  
FT CARBOHYD 627 627 N-linked (GlcNAc...) (Potential).  
FT CONFLICT 14 14 Q -> E (in Ref. 2).  
FT CONFLICT 21 21 V -> L (in Ref. 2).  
FT CONFLICT 487 487 M -> L (in Ref. 2).  
SQ SEQUENCE 648 AA; 74793 MW; 5ADE8F5234F0907E CRC64;

Query Match 29.8%; Score 894.5; DB 1; Length 648;  
Best Local Similarity 36.1%; Pred. No. 1e-50;  
Matches 225; Conservative 98; Mismatches 210; Indels 91; Gaps 20;

QY 1 MVRPQRNKKRPFILILNGVN-----LEVTSKDRPIAVPGSWNE- 38  
DB 27 MLFPKPTSPRELKVLDTGLWSPRADYSNNRLQGFQKQWYRQPLRESGPTLDMVPSSFNDI 86  
QY 39 -QYODLCYEEGPFTYKTFYVPKLSOKHIRLYFAAVNTD---CEVFLNGEKVGENHIEY 94  
DB 87 TQEAELRNFGWVYEREAVALPQRWTDQDRRVLRLINSAAHYAAVWVNGIHVEHEGCH 146  
QY 95 LPFEVDVTGKVGSG---ENELRVVENRLKVGGFPSKVP-----DSGTHVGFFGSF 143  
DB 147 LPFEADITKLVQSGPLTFRTYIAINNTLT---PYTLPGRTIVYKTDPSMPKGYF--V 200  
QY 144 PPAINFDFPYGGIIRPVLIETDHAIRLIDWVDTSESEPEKKLGKVKKIEVSEAVGQE 203  
DB 201 QDISFDFFNYAGLRHRSVVLPTPTTYIDITVTT--DVDRDVGLVNWYISV-QGSDHFQ 256  
QY 204 MTIKLGEEBKIRTSNRFEVEGEFILENARFW---SLEDP-YLYPLKLV-----ELEKDEX 253  
DB 257 LEVRLDDEDEGIIVARGTGNQGLKVPRAHLWMPYLMHEHPAYLYSLEVMTTPESVSDFY 316  
QY 254 TLDIGIRTIISWDEKRLYLNGKPVFLKFGKHEEPVLQGGTFYPLMIKDFNLKMINANS 313  
DB 317 TLPVGIRTVAVTKSKFLINGKPFYHGVNKHEDSDIRGRGFDWPLLIKDFNLLRWLGANS 376  
QY 314 FRTSHYPYSEEWLADRLGLVVIDEAPHVGI--TRYHNPETO---RYHYNPETOKIAEDNIR 363  
DB 377 FRTSHYPYSEVQLCDRYGIVVIDESPGVGLVLPQSFNVSLRHHL-----EVMDELVR 431  
QY 364 RMIDRKNHPSVIMMSVANEPESNHPDAEGFFKALYETANEMDRTRPVVMVSMMDAPDER 423  
DB 432 ----RDKNHPAVVMMSVANEVSSLKRPAGYVFKTLIAHTKALDPTRPVTFVS-----N 480  
CC -----



QY 424 TR--DVALKYFVICVNRYYGYIYQRIEGLQALEKIDIELYARHKKPIFTEFGAD 480  
 DB 481 TRYDADMGAPYVDVICVNSYLSWYHDYGHLEVIQLQTSQFENWYKMYQKPIIOSEYGAD 540  
 QY 481 AIAGIHYDPPQMFSEYQAEIVKTRILL--KDYIIGTHVWAFADFKTPQNVRRPIL 537  
 DB 541 AVSLHEDPPMFSEYQAEIVKTRILL--NYHLIDCKRKEYVIGELIWNFADFTNQSPLRVTG 599  
 QY 538 NHKGVFTDRQPKLVAVHLR-RLW 560  
 DB 600 NKKGIFTRQNRNPKMAAFILRRRYW 623

## RESULT 15

BGLR\_CANFA STANDARD; PRT; 651 AA.  
 AC 018835;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 05-JUL-2004 (Rel. 44, Last annotation update)  
 DE Beta-glucuronidase precursor (EC 3.2.1.31).  
 GN Name=GUSB;  
 OS Canis familiaris (Dog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
 OX NCBI\_TaxID=9615;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND VARIANT MPS VII HIS-166.  
 RX MEDLINE=98190525; PubMed=9521879;  
 RA Ray J., Bouvet A., Desanto C., Fyfe J.C., Xu D., Wolfe J.H.,  
 RA Aguirre G.D., Patterson D.F., Haskins M.E., Henthorn P.S.;  
 RT "Cloning of the canine beta-glucuronidase cDNA, mutation  
 RT identification in canine MPS VII, and retroviral vector-mediated  
 RT correction of MPS VII cells";  
 RL Genomics 48:248-253(1998).  
 CC -!- FUNCTION: Plays an important role in the degradation of dermatan  
 CC and keratan sulfates (By similarity).  
 CC -!- CATALYTIC ACTIVITY: A beta-D-glucuronoside + H(2)O = an alcohol +  
 CC D-glucuronate.  
 CC -!- SUBUNIT: Homotetramer (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Lysosomal.  
 CC -!- DISEASE: Defects in GUSB are the cause of mucopolysaccharidosis  
 CC type VII (MPS VII), an inherited disease reported in humans, mice,  
 CC cats, and dogs.  
 CC -!- SIMILARITY: Belongs to family 2 of glycosyl hydrolases.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF019759; AAC48809.1; -.  
 DR HSSP; P08236; 1BHG.  
 DR InterPro; IPR008979; Gal\_bind\_like.  
 DR InterPro; IPR006101; Glyco\_hydro\_2.  
 DR InterPro; IPR006102; Glyco\_hydro\_2ig.  
 DR InterPro; IPR006104; Glyco\_hydro\_2SB.  
 DR InterPro; IPR006103; Glyco\_hydro\_2TIM.  
 DR Pfam; PF00703; Glyco\_hydro\_2; 1.  
 DR Pfam; PF02836; Glyco\_hydro\_2\_C; 1.  
 DR Pfam; PF02837; Glyco\_hydro\_2\_N; 1.  
 DR PRINTS; PR00132; GLHYDRLASE2.  
 DR PROSITE; PS00719; GLYCOSYL\_HYDROL\_F2\_1; 1.  
 DR PROSITE; PS00608; GLYCOSYL\_HYDROL\_F2\_2; 1.  
 KW Disease mutation; Glycoprotein; Glycosidase; Hydrolase; Lysosome;  
 KW Signal.  
 FT SIGNAL 1 22 By similarity.  
 FT CHAIN 23 651 Beta-glucuronidase.  
 FT ACT SITE 450 450 Proton donor (By similarity).  
 FT CARBOHYD 172 172 N-linked (GlcNAc...) (Potential).

FT CARBOHYD 419 419 N-linked (GlcNAc...) (Potential).  
 FT CARBOHYD 630 630 N-linked (GlcNAc...) (Potential).  
 FT VARIANT 166 166 R -> H (in MPS VII; loss of activity).  
 SQ SEQUENCE 651 AA; 74433 MW; E8991B1E65C60120 CRC64;

Query Match 29.6%; Score 888; DB 1; Length 651;  
 Best Local Similarity 35.7%; Pred. No. 2.8e-50;  
 Matches 220; Conservative 98; Mismatches 226; Indels 72; Gaps 18;

QY 1 MVRPQNRKKRIFILINGWNLEVTGSKD-----RPIAVPGSNNEQ 39  
 DB 27 MLYPRESRSRERKDLGLMSFRADPSDGRQGFEGQWYRAPLRESGPTLDMVPSSFNVDY 86  
 QY 40 YQD--LCYEEGPFYKTFYVPPKXLSQ--KHIRLYFAAVNTDCEVFLNGEKVGENHIEY 94  
 DB 87 GQDRQLRSFVGWVYEREATLPRRRSQDPGTRVLRIGSAHYVALVWNGVHAHEHGH 146  
 QY 95 LPFEVDVTGKVKSG---ENELRVVVENRLKVGFPSPKVP-----DSGTHVGFPGSF 143  
 DB 147 LPFEADISKLVQSGPLSSCRITLAINNTLT---PHTLPRTIYVKTDAKYPKGYF--V 200  
 QY 144 PRANDFPFPYGGIIRPVLIEFTDHARILDIWVDTSESEPEKKLGKVKVIEVSEAVGQE 203  
 DB 201 QNTYFDFPNYAGLHRPVLTYTPTTYIDITVTGVDQ-DTGLVYQIFVQGSSEHF---Q 256  
 QY 204 MTIKLGESEKKIRTSNRFEVEGEFLEENARFW----SLEDP-YLYPLKVELE-----K 250  
 DB 257 LEVYLDEEGKVVAQGTSGRLQVNVHLWVPLMHEHPAYLYSLEVRITQAAGAPVS 316  
 QY 251 DEYTLDIGIRITISWDEKRLYLNGKRVFLKGFKEEFVVLGGTFYPLMIKDFNLKWIN 310  
 DB 317 DFYTLPGVIRTVATERQFLINGKPFYFHGVNKGEDADIRGKGFVPLVKDFNLRLWG 376  
 QY 311 ANSFRTSHYPYSEEWLDLADRLGLIVIDEAPHVGI--TRYHYNPEQKIAEDNIRRMIDR 368  
 DB 377 ANAFRTSHYPYAEVWQLCDRYGIVIDESPGVIMLVQSYNSVSLQHHLE-VMGELVRR 435  
 QY 369 HKNHPVIMSVANEPESNHPDAEGFKALYETANEMDRTRPVWVSMMDAPDERRRDVA 428  
 DB 436 DKNHPSVVMMSVANEPSTFLKPAAYFKTLIAHTKALDPSRPVTFVT---NSNYEADLG 491  
 QY 429 LKYPFIVCVNRYGYIYQRIEGLQALEKIDIELYARHKKPIFTEFGADAIAGIHYD 488  
 DB 492 APYVDVICVNSYLSWYHDYGHMEVIQLQLATEFENWYRTYQKPIIOSEYGAETIAGFHQD 551  
 QY 489 PPQMFSEYQAEIVKTRILL--KDYIIGTHVWAFADFKTPQNVRRPILNHGVFTR 545  
 DB 552 PPLMFSEYQKGLLEQ-YHLVLQDKRKEYVIGELIWNFADFTNQSPQRAVGNRGIFTR 610  
 QY 546 DRQPKLVAVHLR-RLW 560  
 DB 611 QROPKAAAFILRRRYW 626

Search completed: January 24, 2005, 07:52:35  
 Job time : 109 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 24, 2005, 01:17:00 ; Search time 7288 Seconds  
(without alignments)  
10959.427 Million cell updates/sec

Title: US-09-936-759-14  
Perfect score: 1689  
Sequence: 1 atggttaagaccgcaacgaaa.....gaagactgtcgtgaggtc 1689

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4526729 seqs, 23644849745 residues

Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl:\*  
1: gb\_ba:\*  
2: gb\_hcg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pac:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_scs:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1688	99.9	1689	6 AR428149	AR428149 Sequence
2	1688	99.9	12583	1 AE001766	AE001766 Thermotog
3	194	11.5	2153	4 AF012423	AF012423 Felis cat
4	192.4	11.4	2014	4 AF012424	AF012424 Felis cat
5	190.2	11.3	2199	4 AF019759	AF019759 Canis fam
6	174.4	10.3	2155	9 AF084552	AF084552 Chloroceb
7	174	10.3	2128	6 CQ575703	CQ575703 Sequence
8	174	10.3	2182	9 BC014142	BC014142 Homo sapi
9	174	10.3	2321	3 BT010073	BT010073 Drosophil
10	172.4	10.2	1956	6 AR344264	AR344264 Sequence
11	172.4	10.2	1956	6 AX147652	AX147652 Sequence
12	172.4	10.2	2191	9 HDMGLCB	M15182 Human beta-
13	166.8	9.9	2391	10 RNGLUCB	Y00717 Rat mRNA fo
14	166.8	9.9	2472	6 AX827491	AX827491 Sequence
15	166.8	9.9	2472	10 RATGLCB	M13962 Rat beta-gl
16	159.6	9.4	1947	6 AR344265	AR344265 Sequence
17	159.6	9.4	1947	6 AX147654	AX147654 Sequence
18	159.6	9.4	2456	10 MUSGUS	M28540 Mouse beta-
19	159.6	9.4	2456	10 MUSGUS	M28541 Mouse beta-

20	159.6	9.4	2456	10 MUSGUS	M63836 Mouse beta-
21	159.6	9.4	2492	10 BC071226	BC071226 Mus muscu
22	157.8	9.3	2454	10 MUSGLCG	M19279 Mouse beta-
23	156.4	9.3	2455	10 MUSGLC	J03047 Mouse beta-
24	142.2	8.4	23078	6 CQ363720	CQ363720 Sequence
25	142.2	8.4	60265	1 AE017283	Continuation (26 o
26	141.4	8.4	1809	12 AF354047	AF354047 Synthetic
27	141.4	8.4	1887	6 AR210802	AR210802 Sequence
28	141.4	8.4	1887	6 BD074927	BD074927 Microbia
29	141.4	8.4	7018	12 AY452736	AY452736 Reporter
30	141.4	8.4	11846	12 AF354045	AF354045 Binary ve
31	141.4	8.4	11921	12 AF354046	AF354046 Binary ve
32	141.2	8.4	2235	11 BV175016	BV175016 sqm76123
33	139	8.2	10775	1 AE006894	AE006894 Sulfolobu
34	138.6	8.2	2352	3 BT012475	BT012475 Drosophil
35	138.6	8.2	2445	3 AY069329	AY069329 Drosophil
36	130.4	7.7	1888	6 AR428153	AR428153 Sequence
37	127.6	7.6	93590	3 AC084452	AC084452 Caenorhab
38	123.6	7.3	301450	1 AP003185	AP003185 Clostridi
39	122	7.2	3448	1 CPE420784	AJ420784 Clostridi
40	115.6	6.8	2105	1 AF305917	AF305917 Escherich
41	115.6	6.8	300330	1 AP005222	AP005222 Corynebac
42	114	6.7	1814	6 BD185061	BD185061 Nucleic a
43	114	6.7	2103	1 AF305918	AF305918 Escherich
44	114	6.7	10679	1 AE005385	AE005385 Escherich
45	114	6.7	257071	1 AP002558	AP002558 Escherich

ALIGNMENTS

RESULT 1	AR428149	1689 bp	DNA	linear	PAT 18-DEC-2003
LOCUS	AR428149	Sequence 14 from patent US 6641996.			
DEFINITION	AR428149				
ACCESSION	AR428149.1	GI:40187542			
VERSION					
KEYWORDS					
SOURCE	Unknown.				
ORGANISM	Unknown.				
REFERENCE	1 (bases 1 to 1689)				
AUTHORS	Jefferson, R.A. and Mayer, J.E.				
TITLE	Microbial .beta.-glucuronidase genes, gene products and uses thereof				
JOURNAL	Patent: US 6641996-A 14 04-NOV-2003;				
FEATURES	Location/Qualifiers				
source	1. .1689				
ORIGIN	/organism="unknown"				
	/mol_type="genomic DNA"				
Query Match	99.9%;	Score 1688;	DB 6;	Length 1689;	
Best Local Similarity	100.0%;	Pred. No. 0;			
Matches 1689;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;	
QY	1 ATGCTAAGACCCGCAACGAAACAGAGATTATTTCTTATCTTGAATGAGTTGGAAT	60			
Db	1 ATGCTAAGACCCGCAACGAAACAGAGATTATTTCTTATCTTGAATGAGTTGGAAT	60			
QY	61 CTTGAAGTACCAAGACCAACCAATCGCCGTTCTGGAAGCTGGAATGAGCAATAC	120			
Db	61 CTTGAAGTACCAAGACCAACCAATCGCCGTTCTGGAAGCTGGAATGAGCAATAC	120			
QY	121 CAGGATCTGTGCTACGGAAGGAGCCCTTACCTACCAAAACCACTTACGTTCCGAAG	180			
Db	121 CAGGATCTGTGCTACGGAAGGAGCCCTTACCTACCAAAACCACTTACGTTCCGAAG	180			
QY	181 NAACTTACAAAACACATCAGACTTTACTTTGCTGCGGTGAACACGAGCTGCGAGGTC	240			
Db	181 NAACTTACAAAACACATCAGACTTTACTTTGCTGCGGTGAACACGAGCTGCGAGGTC	240			
QY	241 TTCCTCAACGAGAGAAAGTGGAGAGGAATCATTGAATACCTTCCCTTGAAGTAGAT	300			

```
|||||
Db 241 TTCTCAACGAGAGAAAGTGGAGAGATCACATTGAATACCTTCCCTCGAAGTAGAT 300
QY 301 GTGACGGGGAAAAGTGAATCCGAGAGAAAGAACTCAGGGTGGTGTGGAACAGATTG 360
Db 301 GTGACGGGGAAAAGTGAATCCGAGAGAAAGAACTCAGGGTGGTGTGGAACAGATTG 360
QY 361 AAAGTGGAGAGATTTCCCTCGAAGGTTCCAGACAGCGGCACTCACACCGTGGATTGTTT 420
Db 361 AAAGTGGAGAGATTTCCCTCGAAGGTTCCAGACAGCGGCACTCACACCGTGGATTGTTT 420
QY 421 GGAAGTTTCCACCTGCAAACTTGCATCTTCCCTACGGTGAATCATTAAGCCCTGTT 480
Db 421 GGAAGTTTCCACCTGCAAACTTGCATCTTCCCTACGGTGAATCATTAAGCCCTGTT 480
QY 481 CTGATAGATTCCACAGACCGCGAGGATCTCGACATCTGGGTGACACGAGTGAATCT 540
Db 481 CTGATAGATTCCACAGACCGCGAGGATCTCGACATCTGGGTGACACGAGTGAATCT 540
QY 541 GAACCGGAGAGAAACTTGGAAAAGTGAAGTGAAGTGAAGTCTCAGAGAAGCGGTG 600
Db 541 GAACCGGAGAGAAACTTGGAAAAGTGAAGTGAAGTGAAGTCTCAGAGAAGCGGTG 600
QY 601 GGACAGAGATGACGATCAAACTTGAGAGAGAGAGAAAAGATTAGAACATCCACAGA 660
Db 601 GGACAGAGATGACGATCAAACTTGAGAGAGAGAGAAAAGATTAGAACATCCACAGA 660
QY 661 TTCTCGAAGGGGAGTTTCTCTCGAAAACGCCAGTCTGAGCCCTCGAAGATCCATAT 720
Db 661 TTCTCGAAGGGGAGTTTCTCTCGAAAACGCCAGTCTGAGCCCTCGAAGATCCATAT 720
QY 721 CTTTATCTCTCAAGGTGGAATCTTGAAGAGACGAGTACACTCTGACATCGGAATCAGA 780
Db 721 CTTTATCTCTCAAGGTGGAATCTTGAAGAGACGAGTACACTCTGACATCGGAATCAGA 780
QY 781 ACGATCAGCTGGAGACGAGAGAGGCTCTATCTGAACCGGAAAACCTGTTTTGAAGGC 840
Db 781 ACGATCAGCTGGAGACGAGAGAGGCTCTATCTGAACCGGAAAACCTGTTTTGAAGGC 840
QY 841 TTTGAAAGCAGCAGAGAAATCCCGCTTCTGGGCGAGGCACTTTTATCCATTGATGATA 900
Db 841 TTTGAAAGCAGCAGAGAAATCCCGCTTCTGGGCGAGGCACTTTTATCCATTGATGATA 900
QY 901 AAAGACTTCAACCTTCTGAAGTGAATCAACCGGAATCTTTCAGSACCTCTCACTATCT 960
Db 901 AAAGACTTCAACCTTCTGAAGTGAATCAACCGGAATCTTTCAGSACCTCTCACTATCT 960
QY 961 TACAGTGAAGAGTGGCTGGATCTTGGCGACAGACTCGGAATCCTTGTATAGACGAGCC 1020
Db 961 TACAGTGAAGAGTGGCTGGATCTTGGCGACAGACTCGGAATCCTTGTATAGACGAGCC 1020
QY 1021 CCGCAGCTTGTATACAAAGGTACCACTAATCCCGAGACTCAGAAGATAGCAGAGAC 1080
Db 1021 CCGCAGCTTGTATACAAAGGTACCACTAATCCCGAGACTCAGAAGATAGCAGAGAC 1080
QY 1081 AACATAGAAGATGATGACAGACAGACAGAACCAATCCAGTGTATCATGTGAGTGTG 1140
Db 1081 AACATAGAAGATGATGACAGACAGACAGAACCAATCCAGTGTATCATGTGAGTGTG 1140
QY 1141 GCGAAGCAACAGAGTCCAACCATCCAGACGCGGAGGGTTTCTTCAAGCCCTTTATGAG 1200
Db 1141 GCGAAGCAACAGAGTCCAACCATCCAGACGCGGAGGGTTTCTTCAAGCCCTTTATGAG 1200
QY 1201 ACTGCCAATGAATGATCGAACAACCGCCCTGTTGTATGTGAGCATGATGAGCGACCA 1260
Db 1201 ACTGCCAATGAATGATCGAACAACCGCCCTGTTGTATGTGAGCATGATGAGCGACCA 1260
QY 1261 GACGAGAGAACAGAGAGCTGGCGCTGAAGTACTTTCGACATCGTCTGTGTGAACAGGTAC 1320
Db 1261 GACGAGAGAACAGAGAGCTGGCGCTGAAGTACTTTCGACATCGTCTGTGTGAACAGGTAC 1320
QY 1321 TACGGCTGTACATCTATCAGGGAGAGATGAGAGAGGACTTCAAGCTCTGAAAAAGAC 1380
Db 1321 TACGGCTGTACATCTATCAGGGAGAGATGAGAGAGGACTTCAAGCTCTGAAAAAGAC 1380
```

```
Db 1321 TACGGCTGTACATCTATCAGGGAGAGATGAGAGAGGACTTCAAGCTCTGAAAAAGAC 1380
QY 1381 ATAGAAGAGCTCTATGCAAGGCACAGAAAAGCCATCTTTGTACAGAAATTGGTGGGAC 1440
Db 1381 ATAGAAGAGCTCTATGCAAGGCACAGAAAAGCCATCTTTGTACAGAAATTGGTGGGAC 1440
QY 1441 GCCATAGCTGGCATCCACTACGATCCACTCAAAATGTTCTCCGAAGGTACCAAGCAGAG 1500
Db 1441 GCCATAGCTGGCATCCACTACGATCCACTCAAAATGTTCTCCGAAGGTACCAAGCAGAG 1500
QY 1501 CTCGTTGAAAAGAGCAGATCAGGCTCTTTTGAAGAAAAGACTACATCATTCGGAACACAGTG 1560
Db 1501 CTCGTTGAAAAGAGCAGATCAGGCTCTTTTGAAGAAAAGACTACATCATTCGGAACACAGTG 1560
QY 1561 TGGGCTTTGACAGATTTTAAGACTCTCAGAAATGTGAGAAAGCCATTTCTCAACCAAG 1620
Db 1561 TGGGCTTTGACAGATTTTAAGACTCTCAGAAATGTGAGAAAGCCATTTCTCAACCAAG 1620
QY 1621 GGTGTTTTCACAAGAGACAGACAAACCCAACTCGTTGCTCATGTACTGAGAAAGACTGTGG 1680
Db 1621 GGTGTTTTCACAAGAGACAGACAAACCCAACTCGTTGCTCATGTACTGAGAAAGACTGTGG 1680
QY 1681 AGTGAGGTT 1689
Db 1681 AGTGAGGTT 1689

RESULT 2
AE001766/c 12583 bp DNA linear BCT 04-JUN-2004
LOCUS Thermotoga maritima MSB section 78 of 136 of the complete genome.
DEFINITION AE001766 AE000512
ACCESSION AE001766.1 GI:4981600
VERSION
KEYWORDS
SOURCE Thermotoga maritima MSB
ORGANISM Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
REFERENCE 1 (bases 1 to 12583)
AUTHORS Nelson,K.E., Clayton,R.A., Gill,S.R., Gwinn,M.L., Dodson,R.J., Haft,D.H., Hickey,E.K., Peterson,J.D., Nelson,W.C., Ketchum,K.A., McDonald,L., Uterback,T.R., Malek,J.A., Linher,K.D., Garrett,M.M., Stewart,A.M., Cotton,M.D., Pratt,M.S., Phillips,C.A., Richardson,D., Heidelberg,J., Sutton,G.G., Fleischmann,R.D., White,O., Salzberg,S.L., Smith,H.O., Venter,J.C. and Fraser,C.M.
Eisen,J.A., White,O., Salzberg,S.L., Smith,H.O., Venter,J.C. and Fraser,C.M.
EVIDENCE for lateral gene transfer between Archaea and bacteria
from genome sequence of Thermotoga maritima
Nature 399 (6734), 323-329 (1999)
JOURNAL 99287316
MEDLINE 10360571
PUBMED
REFERENCE 2 (bases 1 to 12583)
AUTHORS Nelson,K.E., Clayton,R.A., Gill,S.R., Gwinn,M.L., Dodson,R.J., Haft,D.H., Hickey,E.K., Peterson,J.D., Nelson,W.C., Ketchum,K.A., McDonald,L., Uterback,T.R., Malek,J.A., Linher,K.D., Garrett,M.M., Stewart,A.M., Cotton,M.D., Pratt,M.S., Phillips,C.A., Richardson,D., Heidelberg,J., Sutton,G.G., Fleischmann,R.D., White,O., Salzberg,S.L., Smith,H.O., Venter,J.C. and Fraser,C.M.
Submitted (01-JUN-1999) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
TITLE Direct Submission
JOURNAL
FEATURES
source
1. 12583
/organism="Thermotoga maritima MSB"
/mol_type="genomic DNA"
/strain="MSB"
/db_xref="taxon:243274"
81. .1607
/locus_tag="TM1058"
81. .1607
/locus_tag="TM1058"
/locus_tag="TM1058"
/note="similar to GB:AE000666 percent identity: 48.70; identified by sequence similarity; putative"
/codon_start=1
```

gene /transl\_table=11  
/product="glutamate synthase-related protein"  
/protein\_id="AAD36135.1"  
/db\_xref="GI:4981601"  
/translation="MSGLCACKVDGCTGGCETWLAFRGREVLYPGPFEDITAGAVKD  
YPVDSHLNIIIGYARGAEGLPGEVEPGPDTAIFTNVDTTTEYGMWIKYMKVPIFGA  
LGSTEARKNMDHIAVGAISITVVCENAVGDPDLELDSNGKYKSPELDRIE  
YKRYHDEGEYGEILLQMWEDTRIGVAEYVINKHGITIELKMGQAKSIGEIKVRS  
ERALELKRGRYIVLPPELPEVQRAFEKEIKEFERHSRLGFVSKESEPLKEVERLKL  
GFKRITLKTGAYSAAVELAMALRYGAEKVDLITVDGAPCGTGMSPPMNMENWGPTEY  
LEALTYQPAEKL.SRGRIRVPDIALAGFSTEDVLKAIAMGSPYKAVCMGRALMI.PA  
MVGKNIGEWLKSGLPKTVSKYGTVEEIEFVYEELRSRFGEEVKKLPLGAIGVYTF  
VQKFKTGLQQLMAGARKFRLSLSRKDLIALTKDAEISGIPYMESYRDEAERILIEE  
"  
complement(1650..1913)  
/locus\_tag="TM1059"  
complement(1650..1913)  
/locus\_tag="TM1059"  
/note="similar to SP:P45693 PID:862985 GB:AL009126 percent  
identity: 48.70; identified by sequence similarity;  
putative"  
/codon\_start=1  
/transl\_table=11  
/product="spovs-related protein"  
/protein\_id="AAD36136.1"  
/db\_xref="GI:4981602"  
/translation="MEVLKVSSEKSDPNKVGAGIAGVREHGKAEIQAIAGAVNQAVK  
AIAIARGYLAPSGIDLVPVPAFTVDEIENEKRTAIKFIIVPKPS"  
complement(1962..3272)  
/locus\_tag="TM1060"  
complement(1962..3272)  
/locus\_tag="TM1060"  
/note="similar to GB:AE000782 percent identity: 49.74;  
identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAD36144.1"  
/db\_xref="GI:4981610"  
/translation="WNISIEGALAVLINQFPGAYLTGYFLMWGASSFFIGLFGSIP  
FLANTLQLTL.SFSHRLKRSKQIIVPLMWTARTSILIPAVPAIKGILLAYLYEYI  
QIAGALSAPLMQSWNSDLVPKDMIGSYGFERNLHGAVQIPAMFLAGAILDSLGENWK  
GFGTLPLIAGSLGALNGYFLKIOVEBPYKPREASVSITKAVKFLKEEHFKNLFGRFA  
FMNFAIGVTVYINWMLLKEVEFSYQISVLNAVGMFIGTLFQPFWGLDARYGQYF  
LKVCLMIAHIVILMTLTPRSFLVYFELQIIGIIVTAGTSQLVFTLMTYAPSSLKT  
EAFSVFNSLSNLSLFRAGSLVASVLASLENISLPGISAIRLTMFISFPLRASAAVYII  
SRMDLGPQKVDSLQAVKESFSGTYVPWIRERLNTINIFRRKR"  
complement(3382..4545)  
/locus\_tag="TM1061"  
complement(3382..4545)  
/locus\_tag="TM1061"  
/note="similar to percent identity: 0.00; identified by  
sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAD36137.1"  
/db\_xref="GI:4981603"  
/translation="WRSMTDVRRYWISLKRICEMPLELCASDRLKSMPEVGKSEEB  
RRKFTHELGRIFCGISPLPELNRSPSTDPEERKIATRLSEPAVKSLDVATNPCK  
DYNNFKEGRQPLVDAAFLVEAILRAPKVLWEDLDVSTKKRLIRELKATRKEIPFSNM  
LIFSAMIEETFFPAGSEWDSTKVDLILKNVESWYKGDGAYDGPFRWDYNSFYYP  
MTIDVLRRISEBKEWELVYKVLRAQRYAVVLEWMSPEGTPPIIGRSITYRTAVF  
HLISQSLHLPLPASLPAQVRCALTAVALRIIPENSTFDENGWMLKIGVIGSQPSLGE  
EYITGSLYLCITVFLPLGLPTSDPWRDPCKKTNKKWVEGEDVAPDRALDE"  
complement(4542..6233)  
/locus\_tag="TM1062"  
complement(4542..6233)  
/locus\_tag="TM1062"  
/note="similar to PID:642973 PID:14446 PID:412358 percent  
identity: 57.73; identified by sequence similarity;  
putative"  
/codon\_start=1

gene /transl\_table=11  
/product="beta-glucuronidase"  
/protein\_id="AAD36143.1"  
/db\_xref="GI:4981609"  
/translation="WVRPQRNKRPFILILNGVMNLEVTSKDRPIAVPGSMNEQYDLC  
YEEGFPTYKTFYVPEKLSQKHILRYFAAVNTDCEVFLNGEKEVGENHIEYLPFEVDT  
GKVSGENELRVVENRLKSEVGSKVPDSGTHVGFSGFPAPAFDFPYGGIIRPV  
LIEFTDHARILDIWVDLTSESEPEKLGKVKVIEVSEAVGQEMTIKLGEEBKIRTS  
NRFVEGEFILLENAPWSLEDPLYPLKVELKEDEYTLDIGRITSMDEKRLYLNGRPV  
FLKFGKHEEFVYLGQGTFFPLMKDFNLKMINANSFRTSHYPYSEEWDLADRIGI  
LVIDEAPHVGIITRYHNPETQKIAEDNIRIMIDRKNHPSVIMVAVNEPESNPDAB  
GFFKALYETANEMDRTPVVMVSMMDAPDETRDYALKFEDIVCVNRYGWIYQGR  
IEEGLQALEKDIIEELVYARHKPIFYIEFGADAIAGIHYDPQMFSEBYQAEIVEXTIRL  
LKKDYIIGTHVAFADFKTPQNVRRPILNHKGVFTDRQPKLVAVHLRLMSEV"  
complement(6284..7309)  
/locus\_tag="TM1063"  
complement(6284..7309)  
/locus\_tag="TM1063"  
/note="similar to SP:P42065 PID:677944 GB:AL009126 percent  
identity: 75.31; identified by sequence similarity;  
putative"  
/codon\_start=1  
/transl\_table=11  
/product="oligopeptide ABC transporter, ATP-binding  
protein"  
/protein\_id="AAD36138.1"  
/db\_xref="GI:4981604"  
/translation="MALLEVKNLKKYPIIKRGRRKVVGYLKAVDGISFHIIEGETL  
GLVSESGCGKTTAKLIMRGIEPTEGEIVLNMDEKVDITKSEKELREKGVRRFLQM  
IFQDPYSSLNPRMTVRDIIAEPLVNVGVIKKEEIDAIVSDLLRAVGLRPEYMORYPH  
AFSGGQQRORIAIARAIALKPKLVCDEPTSAIDVSVQAQIINLKDQKEYNLTYLFI  
SHLDGVEHITNRVAVMYVGRIVELAEELFSSRKHPTYEALSAVLPKDPKRRKK  
AQLGVEVPDPSNPPSGCYFHPRCYAKAICKEYVPDLRNVTGDQPHLVACHFADSLK  
LEGVGTM"  
complement(7296..8294)  
/locus\_tag="TM1064"  
complement(7296..8294)  
/locus\_tag="TM1064"  
/note="similar to SP:P42064 PID:677943 GB:AL009126 percent  
identity: 72.95; identified by sequence similarity;  
putative"  
/codon\_start=1  
/transl\_table=11  
/product="oligopeptide ABC transporter, ATP-binding  
protein"  
/protein\_id="AAD36139.1"  
/db\_xref="GI:4981605"  
/translation="MSTLIQIKNRTYFFTDGCVKAVDVSFEIEGRITLGVVSGSG  
CGKSVTARSTIKLSTAGRIIVSGEILLNMDGQWDLVKFSKEEIRKVRGRHIAMTQ  
PMAAFSPVYITIGDQITEGMIYHFGITKQEARERAVELLRVGIPREKMIIDSYPEFS  
GGMQRAMIAMALS CNPRLIADPTTALDVITIAQVLDLKDQOEYKMAIMTTHN  
MGVALEMADHVVMYILGRVESAVEELFYNPKHPTSLLSRISIPVVGKVERLEVIE  
GDVPDPRNMPKGCGRFHPRCPYMMKGI CDEREPEVEVGPEHRVSCFLYGEKDGAS"  
complement(8306..9409)  
/locus\_tag="TM1065"  
complement(8306..9409)  
/locus\_tag="TM1065"  
/note="similar to GB:AB000657 percent identity: 67.80;  
identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="oligopeptide ABC transporter, permease protein"  
/protein\_id="AAD36140.1"  
/db\_xref="GI:4981606"  
/translation="MKKREQMEKEFYVASQWQLIWMFRKHKLAVIGCVLLIIVFA  
IPCEFFAPYDPNKNYRYRYAPRPIHFHEGKEIGPEVGYTYVDLETLRIYKED  
KSKIFKIKFFVRGDEYKFWGIMKTNHIFIGVEDGMFLGTDLSGRDMLSRITYGARI  
STSIGLIGVFLSPILGVAIGIGSYFGGAVDNFIORTIEI IKSIPTIPLMALSAALP  
QNMPLRVYFVYIILSLTGWTDLARVVRSRFLSREDFVMAAKFMGASEARLIFRH  
MLPSEMSHLIASITLSIGMIDETSLSFLGLGRPPIVSWGVVLQEAQNLTVVALYP  
WLIIPVFVITTVLCFNFGDGLDADPYANM"  
complement(9420..10403)  
/locus\_tag="TM1066"

```
CDS
complement(9420..10403)
/locus_tag="TM1066"
/notes="similar to SP:P42062 PID:677946 GB:AL009126 percent
identity: 62.93; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=11
/product="oligopeptide ABC transporter, permease protein"
/protein_id="AAD36141.1"
/db_xref="GI:4981607"

Query Match      99.9%; Score 1688; DB 1; Length 12583;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1688; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGCTAAGACCGCAACGAAAGAGAGATTATCTTATCTTGAATGAGTTTGAAT 60
DB 6233 ATGCTAAGACCGCAACGAAAGAGAGATTATCTTATCTTGAATGAGTTTGAAT 6174
QY 61 CTTGAAGTAACGACAAAGACAGACCAATCCCGCTTCTGGAAGCTGGAATGAGCAGTAC 120
DB 6173 CTTGAAGTAACGACAAAGACAGACCAATCCCGCTTCTGGAAGCTGGAATGAGCAGTAC 6114
QY 121 CAGGATCTGTGCTACGAAAGAGACCCCTTCACTCAAAAACCACTTCTACGTTCCGAAG 180
DB 6113 CAGGATCTGTGCTACGAAAGAGACCCCTTCACTCAAAAACCACTTCTACGTTCCGAAG 6054
QY 181 NAACCTTCACAAAAACACATCAGACTTTACTTTGCTGCGGTGAACACGCACTGCGAGGTC 240
DB 6053 GAACCTTCACAAAAACACATCAGACTTTACTTTGCTGCGGTGAACACGCACTGCGAGGTC 5994
QY 241 TTCCTCAACGAGAGAAAGTGGGAGAGAAATCAATGAAATACCTTCCCTTCGAAGTAGAT 300
DB 5993 TTCCTCAACGAGAGAAAGTGGGAGAGAAATCAATGAAATACCTTCCCTTCGAAGTAGAT 5934
QY 301 GTGACGGGGAAAGTGAATCCGAGAGAGAACTCAGGGTGTGTTGAACAAGATTG 360
DB 5933 GTGACGGGGAAAGTGAATCCGAGAGAGAACTCAGGGTGTGTTGAACAAGATTG 5874
QY 361 AAAGTGGAGGATTTCCCTCGAAGGTTCCAGACAGCGGCACTCACACCGTGGGATTTT 420
DB 5873 AAAGTGGAGGATTTCCCTCGAAGGTTCCAGACAGCGGCACTCACACCGTGGGATTTT 5814
QY 421 GGAAGTTTCCACCTGCAAACTTCGACTTCTCCCTACGGTGAATCATTAAGCCCTGTT 480
DB 5813 GGAAGTTTCCACCTGCAAACTTCGACTTCTCCCTACGGTGAATCATTAAGCCCTGTT 5754
QY 481 CTGATAGATTACAGACCAAGGAGATCTCGACATCTGGGTGACACGAGTAGTCT 540
DB 5753 CTGATAGATTACAGACCAAGGAGATCTCGACATCTGGGTGACACGAGTAGTCT 5694
QY 541 GAACCGGAGAAAGAACTTGGAAAAGTGAAGATGAAGTCTCAGAAAGACCGGTG 600
DB 5693 GAACCGGAGAAAGAACTTGGAAAAGTGAAGATGAAGTCTCAGAAAGACCGGTG 5634
QY 601 GGACAGAGATGACGATCAAACTTGAGAGAGAGAAAAGATTAGAACATCCAACAGA 660
DB 5633 GGACAGAGATGACGATCAAACTTGAGAGAGAGAAAAGATTAGAACATCCAACAGA 5574
QY 661 TTCTGCAAGGGGAGTTTCTCTCGAAAACGCCAGGTTCTGAGCCCTGAAAGATCCATAT 720
DB 5573 TTCTGCAAGGGGAGTTTCTCTCGAAAACGCCAGGTTCTGAGCCCTGAAAGATCCATAT 5514
QY 721 CTTTATCTCTCAAGGTGAAGTGAAGAAAGACGATACATCTTGACATCGGAATCAGA 780
DB 5513 CTTTATCTCTCAAGGTGAAGTGAAGAAAGACGATACATCTTGACATCGGAATCAGA 5454
QY 781 ACGATCAGCTGGGACGAGAAAGAGGCTTATCTGAACGGGAAACCTGTCTTTTGAAGGCG 840
DB 5453 ACGATCAGCTGGGACGAGAAAGAGGCTTATCTGAACGGGAAACCTGTCTTTTGAAGGCG 5394
QY 841 TTTGAAAGACGAGGAATTCCTGCTTGGGACAGGCACTTTTATCCATTGATGATA 900
DB 841 TTTGAAAGACGAGGAATTCCTGCTTGGGACAGGCACTTTTATCCATTGATGATA 900
```

```
DB 5393 TTTGAAAGACGAGGAATTCCTGCTTCTGGGACAGGCACTTTTATCCATTGATGATA 5334
QY 901 AAAGACTTCAACCTTCTGAAGTGAATCAACGCGAATTTCTTACAGACCTCTACATACCT 960
DB 5333 AAAGACTTCAACCTTCTGAAGTGAATCAACGCGAATTTCTTACAGACCTCTACATACCT 5274
QY 961 TACAGTGAAGAGTGGCTGATCTTGGCCGACAGACTCGGAATCTTGTATAGACGAAGCC 1020
DB 5273 TACAGTGAAGAGTGGCTGATCTTGGCCGACAGACTCGGAATCTTGTATAGACGAAGCC 5214
QY 1021 CCGACGTTGATACCAAGGTACCACTACCAATCCCGAGACTCGAAAGATAGCAGAGAGAC 1080
DB 5213 CCGACGTTGATACCAAGGTACCACTACCAATCCCGAGACTCGAAAGATAGCAGAGAGAC 5154
QY 1081 AACATTAAGAAATGATCGACAGACACAAGAACCAATCCAGTGTGATCATGTGAGTGTG 1140
DB 5153 AACATTAAGAAATGATCGACAGACACAAGAACCAATCCAGTGTGATCATGTGAGTGTG 5094
QY 1141 GCGAAGAACCAAGAGTCCAAACCAATCCAGACGCGAGGGTTTCTTCAAGCCCTTATGAG 1200
DB 5093 GCGAAGAACCAAGAGTCCAAACCAATCCAGACGCGAGGGTTTCTTCAAGCCCTTATGAG 5034
QY 1201 ACTGCCAATGAATGATCGAACACAGCCCGCTTGTATGTGATGATGATGACGACCA 1260
DB 5033 ACTGCCAATGAATGATCGAACACAGCCCGCTTGTATGTGATGATGATGACGACCA 4974
QY 1261 GACGAGAGAACCAAGAGAGCGTGGCGCTGAAGTACTTGCACATCGTCTGTGAACAGGTAC 1320
DB 4973 GACGAGAGAACCAAGAGAGCGTGGCGCTGAAGTACTTGCACATCGTCTGTGAACAGGTAC 4914
QY 1321 TACGGCTGTACATCTATCAGGGAAGATAGAAAGACTTCAAGCTCTGGAAGAAAGAC 1380
DB 4913 TACGGCTGTACATCTATCAGGGAAGATAGAAAGACTTCAAGCTCTGGAAGAAAGAC 4854
QY 1381 ATAGAAGAGCTTATGCAAGGACAGAAAGCCCATCTTGTCAAGAAATCGGTCGGGAC 1440
DB 4853 ATAGAAGAGCTTATGCAAGGACAGAAAGCCCATCTTGTCAAGAAATCGGTCGGGAC 4794
QY 1441 GCGATAGCTGGCATCCACTACGATCCACCTCAAAATGTTCTCGAAAGATACCAAGCAGAG 1500
DB 4793 GCGATAGCTGGCATCCACTACGATCCACCTCAAAATGTTCTCGAAAGATACCAAGCAGAG 4734
QY 1501 CTGTTGAAAAGACGATCAGGCTCTTTGAAAAAAGAACTACATCATCGGAACACAGTG 1560
DB 4733 CTGTTGAAAAGACGATCAGGCTCTTTGAAAAAAGAACTACATCATCGGAACACAGTG 4674
QY 1561 TGGGCTTTGACAGATTTTAAAGCTCCTCAGAAATGTGAGAAAGCCATTCTCAACCAAG 1620
DB 4673 TGGGCTTTGACAGATTTTAAAGCTCCTCAGAAATGTGAGAAAGCCATTCTCAACCAAG 4614
QY 1621 GGTGTTTTCACAAGACAGACAAACCAAACTCGTTGCTCATGTACTGAGAAAGCTGTGG 1680
DB 4613 GGTGTTTTCACAAGACAGACAAACCAAACTCGTTGCTCATGTACTGAGAAAGCTGTGG 4554
QY 1681 AGTGAGGTT 1689
DB 4553 AGTGAGGTT 4545
```

```
RESULT 3
AF012423 2153 bp mRNA linear MAM 09-SEP-1999
LOCUS AF012423
DEFINITION Felis catus beta-glucuronidase (GUSB) mRNA, complete cds.
ACCESSION AF012423
VERSION AF012423.1 GI:4102550
KEYWORDS
SOURCE
ORGANISM Felis catus (cat)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.
REFERENCE
1 (bases 1 to 2153)
Fyfe,J.C., Kurzhaas,R.L., Lassaline,M.E., Henthorn,P.S., Alur,P.R.,
Wang,P., Wolfe,J.H., Giger,U., Haskins,M.E., Patterson,D.F.,
```



TITLE  
Sun,H., Jain,S. and Yuhki,N.  
Molecular basis of feline beta-glucuronidase deficiency: an animal  
model of mucopolysaccharidosis VII  
JOURNAL  
Genomics 58 (2), 121-128 (1999)  
MEDLINE  
99296826  
PUBMED  
10366443  
REFERENCE  
2 (bases 1 to 2153)  
AUTHORS  
Pyfe,J.C., Kurzhals,R.L. and Henthorn,P.S.  
TITLE  
Direct Submission  
JOURNAL  
Submitted (03-JUL-1997) Microbiology, Michigan State University,  
413 Giltner Hall, East Lansing, MI 48824, USA  
FEATURES  
source  
1. .2153  
/organism="Felis catus"  
/mol\_type="mRNA"  
/db\_xref="taxon:9685"  
/feature\_type="liver"  
1. .2153  
/gene="GUSB"  
24. .1979  
/gene="GUSB"  
/EC\_number="3.2.1.31"  
/note="lysosomal hydrolase; lysosomal exooligosaccharidase  
in stepwise degradation of glucuronic acid-containing  
oligosaccharide sidechains of proteoglycans"  
/codon\_start=1  
/product="beta-glucuronidase"  
/protein\_id="AAB01498.1"  
/db\_xref="GI:4102551"  
/translation="MARGPAVWALGPLMACGLARGMLYPRESPSRERKELNGL  
WSPRADFSENRRQFEQWYRTPLRESGPTLDMVPVSSFNDVGODROLRSFVGWWE  
REATLPQRTQDLGTRVLRIGSAHYAIWVNGVVAHEGHLPEADISKLVQSG  
PLASCRITTAIINNTLTPHTLPSTILYQDTDSKYPKGYFVQNIQFENYAGLHREVL  
LYTPPTYIDIDITISTSVNODTGLVDYQIFVEGGEHQLEVRLLDEGKVVAQGTGR  
GOLQVFNALMWPYLMHEHPAYLSLEVRLLTAQTAGSVSDFTYLPVIGIRTAATVHQ  
FLINGKPFYFHGVNKEDADIRGKFDWPLVKDNLRLMGLANARTSHYPYAEVM  
QLCDRYGIVVIDESPVGIVLVESYSNVSLOHLEVMELVLRDKNHPAVMMSVANE  
PASFLKPAGYFFKTLIAHTKALDPSRPVTFVTSNYEADLGAPYDVICVNSYSMYH  
DYHMEVITQLATQFENMYRTYOKPILIOSEYADTIAGFHODPLMSESEYOKGLE  
QYHLVLDQKRKEYVVGELIWNFADPMTNOSPQVWGNKKGIPTRRQPKGAFLIBER  
YMKLANETRYPMSAVYSQCLENSPFTL"  
ORIGIN  
Query Match 11.5%; Score 194; DB 4; Length 2153;  
Best Local Similarity 53.1%; Pred. No. 7.3e-41;  
Matches 494; Conservative 0; Mismatches 415; Indels 21; Gaps 3;  
Qy 751 GACGAGTACACTCTGGACATCGAATCAGAACGATCAGCTGGGACGAGAAGGCTCTAT 810  
Db 972 GACTTCTACACTCTCCCGTGGGATTCGACGCTGGCTGTCAAGACGACCAAGTTCTC 1031  
Qy 811 CTGAACGGGAAACCTGCTTTTGAAGGCTTTGAAAGCAGACGAGAATCCCGTTCTG 870  
Db 1032 ATCAATGGGAAACCTTCTATTCTCACGGGGTCAACAAGCAGAGATGACATATCCGA 1091  
Qy 871 GGGCAGGGCACCTTTATCCATGTGATATAAAGACTTCAACCTTCTGAAGTGATCAAC 930  
Db 1092 GGGAAAGGCTTTGACTGGCCACTGCTGTGAAGACTTCAATTGCTTGGCTCGGG 1151  
Qy 931 GCGAATCTTTTCAGGACCTCTCACTATCTTACAGTGAAGATGGCTGATCTGCCGAC 990  
Db 1152 GCCAAGCGCTTCCGACACGACTACCCCTACGCGAGAGGTGATGACACTCTGTGAC 1211  
Qy 991 AGACTCGGAATCCTGTGATAGCAGAACCCCGCACGTTGTGATACAAAGTACCA--C 1047  
Db 1212 CGCTATGGGATCGTGTATCATGACGAGATCCCGGTGTGGCATCGTGTGTGAGAGC 1271  
Qy 1048 TACAATCCCGAGACTCAGAAAGATAGCAGAACCAACATTAAGAAATGATGACAGACAC 1107  
Db 1272 TACAGCAACGTGTCTCTGACAGACCACTGTGAGGTGATGAGAGCTGTGCGCAGGAC 1331  
Qy 1108 AAGAACCATCCAGTGTGATCATGTGAGTGTGCGAAGCAACGACGAGTCCAAACATCCA 1167

Db 1332 AAGAATCACCCGGCTGTTGATGTGCTGTGGCCAAATGAGCCCGCTTCTCTGAAA 1391  
Qy 1168 GACCGGAGGCTTCTTCAAAAGCCCTTTATGAGACTGCCAATGAATGATGAAACAGC 1227  
Db 1392 CTGCGCGTACTACTTCAAGACGCTGATGTGCTACACCAAGCCTTGATCCTCCCGG 1451  
Qy 1228 CCCGTTGTATGTGAGCATGATGAGCAGCAGACGACGAGAACAAAGACGTGGCGCTG 1287  
Db 1452 CCCGTGACCTTTGTGACCAACTCCAACTATGAAG-----CAGACCTGGGGCG 1499  
Qy 1288 AAGTACTTCGACATCGTCTGTGTGAACAGGACTACGCGCTGTATCATCTATCAGGAAG 1347  
Db 1500 CCGTATGTGACGTCACTGTGTGAATAGTTACTACTCTTGTGTATCATGACTATGTGTCAC 1559  
Qy 1348 ATAGAAGAAGACTTCAAGCTCTGAAAAAGACATAGAAGACTTATGCAAGGACAGAGA 1407  
Db 1560 ATGAGGTGATTCAGCTGCAGCTGGCAACCCAGTTTGAGAACTGTATAGACCTATCAG 1619  
Qy 1408 AAGCCATCTTTGTACACAGAAATTCGGTCCGACGCGATAGCTGCATCCACTACGATCCA 1467  
Db 1620 AAACCAATTAATCCAGAGCGAGTACGAGCAGACACCATTTGAGAACTTGACAGACCCA 1679  
Qy 1468 CCTCAATGTTCTCCGAAGATACCAAGCAGAGCTCGTTGAAAAGACGATCAGGCTCCTT 1527  
Db 1680 CCTGTGATGTTACGAGGAATATCCAGAAAGGTCTGCTGAGCAGTATCATTTGTTCTG 1739  
Qy 1528 TTGAAA-----AAAGACTCATCATCGGAACACACAGTGTGGCCTTTGACATTTTAAG 1581  
Db 1740 GATCAAAAACGCAAAAGATAGTGTGCGAGAGCTCATCTGAACTTTGCCATTTTATG 1799  
Qy 1582 ACTCCTCAGAAATGTGAGAAGACCCATTCTCAACCAACAAGGCTTTTCAAGAAGACAGA 1641  
Db 1800 ACTAACCAAGTACACCGCAGAGAGTAAATGGGATAAAAAAGGGAATCTTACGCGCCAGAGA 1859  
Qy 1642 CAACCCAAACTGTTGCTCATGTACTGAGA 1671  
Db 1860 CAACCAAAAGGGCAGCGTCTCTTTTGGCA 1889  
RESULT 4  
AF012424  
LOCUS 2014 bp mRNA linear MAM 09-SEP-1999  
DEFINITION Felis catus cell-type fibroblast mutant beta-glucuronidase (GUSB)  
ACCESSION AF012424  
VERSION AF012424.1 GI:4102552  
KEYWORDS  
SOURCE  
ORGANISM  
Felis catus (cat)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.  
REFERENCE  
1 (bases 1 to 2014)  
Pyfe,J.C., Kurzhals,R.L., Lassaline,M.E., Henthorn,P.S., Alur,P.R.,  
Wang,P., Wolfe,J.H., Giger,U., Haskins,M.E., Patterson,D.F.,  
Sun,H., Jain,S. and Yuhki,N.  
TITLE  
Molecular basis of feline beta-glucuronidase deficiency: an animal  
model of mucopolysaccharidosis VII  
JOURNAL  
Genomics 58 (2), 121-128 (1999)  
MEDLINE  
99296826  
PUBMED  
10366443  
REFERENCE  
2 (bases 1 to 2014)  
AUTHORS  
Pyfe,J.C., Kurzhals,R.L. and Henthorn,P.S.  
TITLE  
Direct Submission  
JOURNAL  
Submitted (03-JUL-1997) Microbiology, Michigan State University,  
413 Giltner Hall, East Lansing, MI 48824, USA  
FEATURES  
source  
1. .2014  
/organism="Felis catus"  
/mol\_type="mRNA"  
/db\_xref="taxon:9685"  
/cell\_type="fibroblast"  
1. .2014  
/gene="GUSB"

CDS

/allele="MPS VII mutant"  
2.1957  
/gene="GUSB"  
/allele="MPS VII mutant"  
/EC\_number="3.2.1.31"  
/codon\_start=1  
/product="mutant beta-glucuronidase"  
/protein\_id="AA01499.1"  
/translation="MLRGPAAVMAALGPLLMACGLALRGMLYPRESRERKELNGL  
MSFRADFSRNRQGFEOQWYRPLRESGPTLDMVPSSFNDVGQDRLRSFVGWVWE  
REATLPRWSDGTRVLRIGSAHYAIVWNGVHVAHEGHLPEADISKLVOGS  
PLASCRITIAINNTLPHTLPPGTILYQDTSKYPKGYFVQNTNEDFENYAGLHPVL  
LYTPTTYIDITITSVNODTGLVDYQIFVEGGEHQLVRLDEGKVAQGTGR  
GLOVPNHLMPYLMHEHPAYLSLEVRLTAQTAAGSVSDFTLLPVGIRIVAVTERQ  
FLINGKPFYFHVGNKHEDADIRKGFDMPLVKDFNLRLWLGANAFRTSHYPYAEVM  
QLCDRYGIVVIDESPGVIGLVESYSNVSLOHLEVMELVRDKNHPAVMMSVANE  
PASFLKPAAYYFKTLIAHTKALDPSRPVTFVNSNYEADLGAAPYDVICVNSYYSWYH  
DYGHEVIOQLATPEFNWRYQKPIIOSEYGAETIAGFHQDPLMPSSEYQKGLLE  
OYHLVLDQKREYVVGELIWNFADFMFNQSPQRAVGNRKGIFFTRQRPKAAPFLLRER  
YMKLANETRYPWSAVKSQCLENSPFTL"

ORIGIN

Query Match 11.4%; Score 192.4; DB 4; Length 2014;  
Best Local Similarity 53.0%; Pred. No. 2e-40;  
Matches 493; Conservative 0; Mismatches 416; Indels 21; Gaps 3;

Qy 751 GACGAGTACACTCTGACATCGGAATCAGACGATCAGCTGGAGCAGAGAAGGCTCTAT 810  
Db 950 GACTTCTACACTCTCCCGTGGGAGATTGCGACGGTGTCTCAGACAGACCAAGTTCCCTC 1009  
Qy 811 CTGAACGGGAAACCTGCTTTTGAAGGCTTTGAAAGCAGAGAATTCCTGCTG 870  
Db 1010 ATCAATGGGAAACCTTCTATTTCCACGGGGTCAACAAGCACAAGATGAGATATCCGA 1069  
Qy 871 GGGCAGGGCACTTTTATCCATGTGATGATAAAGACTTCACTTCTGAAGTGAACAAC 930  
Db 1070 GGGAAAGGCTTTGACTGGCCACTGTGTGAAGACTTCAATTTGCTGCTGCTCGG 1129  
Qy 931 GCGAATCTTTGAGGACCTCTCACTATCTTACAGTGAAGAGTGGCTGATCTTCCGAC 990  
Db 1130 GCCAAGCCTTTCCGACCACTACCTTACGCGAGAGGTGATGACGCTCTGTGAC 1189  
Qy 991 AGACTCGGAATCCTTGTGATAGCAGAACCCCGCACGTTGGTATCACAAGGTACCA---C 1047  
Db 1190 CGCTATGGATCGTGTGATCGACAGAGATCCCGGTGGGCATCGTGTGTCAGAGC 1249  
Qy 1048 TACAATCCGAGACTCAGAGATAGCAGAGACAACATAGAAGAATGATCGACAGACAC 1107  
Db 1250 TACAGCAACGTGTCTGACAGACCACTGGAAGTGAAGAGAGTGTGCGCAGGAC 1309  
Qy 1108 AAGAACCATCCAGTGTGATGATGTGGCGAAGCAGAGTCCCAACCATCCA 1167  
Db 1310 AAGAATCACCCGGCTGTGTGATGTGTGTGGCCAAATGAGCCCGCTTCTCTGAAA 1369  
Qy 1168 GACGGGAGGGTTCTTCAAGCCCTTTATGAGACTGCCAATGAATGATCGAAGCAGC 1227  
Db 1370 CCTGCCGTTACTACTTCAAGACGCTGATGCTCACACAAAGCCTTGATCCCTCCGG 1429  
Qy 1228 CCCGTTGTCATGTGAGCATGATGACGCAACAGACGAGAACAGACGAGCGCTG 1287  
Db 1430 CCCGTGACCTTTGTGACCAACTCCAACTATGAAG-----CAGACCTGGGGGG 1477  
Qy 1288 AAGTACTTGCACATCGTGTGTGAACAGTACTACGGCTGTACATCTATCAGGGAAG 1347  
Db 1478 CCGTATGTGACGTCATCTGTGTGAATAGTACTCTTGTGATCATGACTATGTGTAC 1537  
Qy 1348 ATAGAAGAAGACTTCAAGCTCTGAAAAAGACATAGAAGAGCTTATGCAAGCAGACA 1407  
Db 1538 ATGAGGTGATTCAGCTGCAAGCTGCAACCCAGTTGAGAACTGTATAGACCTATCAG 1597  
Qy 1408 AAGCCCATCTTTGTACAGAAATTCGTCGGAGCGGATAGCTGCACTACAGATCCA 1467

Db 1598 AAACCAATATCCAGAGCGAGTAGCGAGACAGACCACTTGCAAGGTTTACCAGAGCCA 1657  
Qy 1468 CCTCAATGTTCTCCGAAAGAGTACCAAGCAGAGCTCGTTGAAAAGCATCAGGCTCTT 1527  
Db 1658 CCTGTATGTTTACGAGGAGATACCAAGAGTCTGCTGAGCAGTATCATTTGTTCTG 1717  
Qy 1528 TTGAA-----AAGACTACATCATCGGAACACACGCTGGGCTTTGCAGATTTTAAG 1581  
Db 1718 GATCAAAAGCAAGAAATACGTGTGAGAGGCTCATCTGAACTTTGCCGATTTATG 1777  
Qy 1582 ACTCTCAGATGTGAGAGACCACTTCTCAACCACAAGGCTGTTTCAACAAGACAGACA 1641  
Db 1778 ACTAACAGTACACCGCAGAGAGTAAAGGAAATAAAGGAGATCTTCAACGCGCAGAGA 1837  
Qy 1642 CAACCAAACTCGTCTCATGTACTGAGA 1671  
Db 1838 CAACCAAAAGGCGAGCGCTTCTTTTGGCA 1867

RESULT 5  
AF019759 2199 bp mRNA linear MAM 30-MAR-1998

LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

AF019759 2199 bp mRNA linear MAM 30-MAR-1998  
Canis familiaris beta-glucuronidase (GUSB) mRNA, complete cds.  
AF019759  
AF019759.1 GI:2425090

REFERENCE  
AUTHORS  
TITLE

1 (bases 1 to 2199)  
Ray, J., Bouvet, A., Desanto, C., Fyfe, J.C., Xu, D., Wolfe, J.H.,  
Aguiar, G.D., Patterson, D.F., Haskins, M.E. and Henthorn, P.S.  
Cloning of the canine beta-glucuronidase cDNA, mutation  
identification in canine MPS VII, and retroviral vector-mediated  
correction of MPS VII cells  
Genomics 48 (2), 248-253 (1998)

JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

9521879  
2 (bases 1 to 2199)  
Henthorn, P.S., Fyfe, J.C. and Bouvet, A.B.  
Direct Submission  
Submitted (15-AUG-1997) Section of Medical Genetics, University of  
Pennsylvania School of Veterinary Medicine, 3900 Delancey St.,  
Philadelphia, PA 19104, USA

FEATURES  
source  
location/Qualifiers

1..2199  
/organism="Canis familiaris"  
/mol\_type="mRNA"  
/db\_xref="taxon:9615"  
1..2199  
/gene="GUSB"  
63..2018  
/gene="GUSB"  
/EC\_number="3.2.1.31"  
/note="lysosomal enzyme"  
/codon\_start=1  
/product="beta-glucuronidase"  
/protein\_id="AA04809.1"  
/db\_xref="GI:2425091"

gene  
CDS

1..2199  
/translation="MSRGPAGAWVALGPLMTGALLEGMLYPRESRERKDLGL  
MSFRADFSRNRQGFEOQWYRPLRESGPTLDMVPSSFNDVGQDRLRSFVGWVWE  
REATLPRWSDGTRVLRIGSAHYAIVWNGVHVAHEGHLPEADISKLVOGS  
PLASCRITIAINNTLPHTLPPGTILYQDTSKYPKGYFVQNTNEDFENYAGLHPVL  
LYTPTTYIDITITSVNODTGLVDYQIFVEGGEHQLVRLDEGKVAQGTGR  
GLOVPNHLMPYLMHEHPAYLSLEVRLTAQTAAGSVSDFTLLPVGIRIVAVTERQ  
FLINGKPFYFHVGNKHEDADIRKGFDMPLVKDFNLRLWLGANAFRTSHYPYAEVM  
QLCDRYGIVVIDESPGVIGLVESYSNVSLOHLEVMELVRDKNHPAVMMSVANE  
PASFLKPAAYYFKTLIAHTKALDPSRPVTFVNSNYEADLGAAPYDVICVNSYYSWYH  
DYGHEVIOQLATPEFNWRYQKPIIOSEYGAETIAGFHQDPLMPSSEYQKGLLE  
OYHLVLDQKREYVVGELIWNFADFMFNQSPQRAVGNRKGIFFTRQRPKAAPFLLRER  
YMKLANETGHRHSAKSQCLENSPFAL"

ORIGIN

Query Match		11.3%;	Score 190.2;	DB 4;	Length 2199;
Best Local Similarity		52.8%;	Pred. No. 7.7e-40;		
Matches 492;		Conservative 0;	Mismatches 418;	Indels 21;	Gaps 3;
QY	750	AGACGAGTACACTCTGGACATCGGAATCAGAACGATCAGCTGGACGAGAAAGGCTCTA	809		
DB	1010	AGACTTCTATACTCTCCCGTGGGATTCGCACCGTGGCCGTCAAGAGCGCCAGTTCTT	1069		
QY	810	TCTGAACGGAAACCTGTCTTTTGAAGGCTTTGAAAGCAGAGGAATTCCTCCCTTCT	869		
DB	1070	CATCAACGGAAACCTTTCTATTCCATGGGGTCAACAACATGAGATGCCGATATCCG	1129		
QY	870	GGGGCAGGGCACCCTTTTATCCATGTATGATAAAGACTTCAACCTTCTGAAGTGATCAA	929		
DB	1130	AGGGAAGGGCTTTGACTGGCCGCTGTGTGAAGACTTCAACCTGTGCGCTGCTGGG	1189		
QY	930	CGGGAATCTTTGAGGACCTCTCACTATCCTTACAGTGAAGAGTGCTGGATCTTGCCGA	989		
DB	1190	CGCCATGCTTCCGACCAAGCCACTACCCCTACCGGAGAGAGTGATGCAAGCTTCCGA	1249		
QY	990	CAGACTCGGAATCCTTGTGTATAGACGAAGCCCCGCACTTGTGTATCAAGGTACCA---	1046		
DB	1250	CCGCTATGGGATGCTGTGTCATCGACGAGAGCCCTGTGTGTGGCATCATGCTGCTCAGAG	1309		
QY	1047	CTACCAATCCCGAGACTCAGAAAGATAGCAGAAAGACAATAGAGAATGATCGACAGACA	1106		
DB	1310	CTACAGCAATGTGTCTCTGACGACCATCTGGAGTGATGGGGAGCTGTGCTCGGGA	1369		
QY	1107	CAAGAACCATCCCAAGTGTGATCATGTGAGTGTGGCGGAACGAAACAGAGTCCCAACATCC	1166		
DB	1370	TAAGAATCACCCATCTGTAGTCACTGTGTCTGTAGCCCAATGAGCCCACTTCTTCTGAA	1429		
QY	1167	AGACGGGAGGGTTTCTCAAAAGCCCTTATGAGACTGCCAATGAATGATCGAACACG	1226		
DB	1430	GCCTGCTGCTTACTACTTCAAGACGCTGATGCTCAACACCAAGGCTTGAGCCCTCCCG	1489		
QY	1227	CCCCGTTGTCATGTGAGCATGATGAGCAGCAGCAGAGAGAAACAAGAGCGTGGCGCT	1286		
DB	1490	GCCCGTGACCTTTGTGACCAATTCCAATATGAG-----CAGACCTGGGGGC	1537		
QY	1287	GAACTACTTCGACATCGTCTGTGTGAACAAGGTACTACGGCTGTATCATCTATCAGGGAG	1346		
DB	1538	GCCGTATGTGAGCTCATCTGTGTCAACAGTTACTCTTGTATCAGCAGTATGGGCA	1597		
QY	1347	GATAGAGAAGAAGCTTCAAGCTTGGAAAAAAGCATAGAAAGCTCTATGCAAGGCCACAG	1406		
DB	1598	CATGAGGTGATTCACTGCTGAGCTGGCCACCGAGTTTGAAGCTGGTATAGGACTTACCA	1657		
QY	1407	AAAGCCCATCTTTGTACAGAAATTCGGTCCGACGCGAGTAGCTGGCATCCACTACGATCC	1466		
DB	1658	GAAACCAATATCCAGAGCGAGTACGGGGCAGAGACAATTGCAAGGCTTCCACAGGATCC	1717		
QY	1467	ACCTCAATGTCTCTCGAAGAGTACCAAGCAGAGCTGTTGAAAAGAGCATCAGGCTCCT	1526		
DB	1718	ACCTGTATGTTCAGTGAAGAGTACAGAAAGGTCTGCTGACAGATATCATTTGTGCT	1777		
QY	1527	TTTGA-----AAAGACTATCATTCGGAACACACGTTGGGCTTTGAGATTTTAA	1580		
DB	1778	GGATCAGAAACGAAAGATATGTGTGAGAGCTCATCTGGAATTTTGTGATTTTAT	1837		
QY	1581	GACTCTCAGATGTGAGAAGACCCATTTCTCAACCAAGAGGTGTTTTCACAGAGACAG	1640		
DB	1838	GACTGACAGTCAACACAGAGAGCAGTAGGGAAGAAAGGCACTTCACTCGCCAGAG	1897		
QY	1641	ACAACCCAAACTGTTGCTCATGTACTGAGA	1671		
DB	1898	ACAACCCAAAGCGCGGCTTCTTTTGGCA	1928		

RESULT 6  
AF084552  
LOCUS AF084552 2155 bp mRNA linear PRI 07-SEP-1998

DEFINITION	Chlorocebus aethiops beta-glucuronidase mRNA, partial cds.
ACCESSION	AF084552
VERSION	AF084552.1 GI:3549608
KEYWORDS	
SOURCE	
ORGANISM	Cercopithecus aethiops (African green monkey)
REFERENCE	Cercopithecus aethiops
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae; Cercopithecinae; Cercopithecus.
TITLE	1 (bases 1 to 2155)
JOURNAL	Vervoot R.
REFERENCE	Partial cDNA sequence of Cercopithecus aethiops (COS7 cell)
AUTHORS	beta-glucuronidase
TITLE	Unpublished
JOURNAL	2 (bases 1 to 2155)
JOURNAL	Vervoot R.
JOURNAL	Direct Submission
JOURNAL	Submitted (16-AUG-1998) Cell Genetics, Medical Research Council Human Genetics Unit, Crewe Road, Edinburgh EH4 2XU, U.K.
FEATURES	Location/Qualifiers
source	1. .2155
	/organism="Cercopithecus aethiops"
	/mol_type="mRNA"
	/db_xref="taxon:9534"
	/cell_line="COS7"
	<1. .1948
	/EC_number="3.2.1.31"
	/function="hydrolase"
	/note="similar to Homo sapiens beta-glucuronidase; beta-D-glucuronoside; glucuronosohydrolase; glucuronidase; glucuronohydrolase"
	/codon_start=2
	/product="beta-glucuronidase"
	/protein_id="AAC34593.1"
	/db_xref="GI:3549609"
	/translation="GLAMAWAVLGPDLWGALALQGMLYPRESQSRERKELDLMSF RADFSNRRRGFEQWYRRPLRESGPTLDMVPSSPNDISQWRLRHFVGWVYREYV ILPERWODLSTRVLRIGSAHAYAIWVNGVHLEGGYLPEDDISNLVQVPLS SHVITIAINNTLTSTLPQTYLTDISKYPRGYIQTNYFDENYAGLQSLVLY TTPTAYIDITVTVEHDTGLVYQISVKSNIPELEVLRLDANKLVANGTIGCQ LKVPBARLWMPYLMHERPAYLYSLEVRLTAGTSGLPVSDFYTLPGIRTVAVTESQFL INKPYFHGVNKHEDADIRKGFDPVLVKDFNLRLWLGANAFRTSHYPYAEVLQ CDRYGIVIDECPVGLALPQFENNVSQNHMRWMEVVRDKNRPVAVMWSVANBA SHLESAGYLLKMYLTHTKALDPSRPTVFTNSNYADKGAAYDVICLSYYSWYDY GHELELIQRLTTOFENWYKTYQKPIIQSEYGAETIVGFHQDPPLMFTBEYQKSLLEQY HVVLDOQRKRYVVGELIWNFADFMTEQSPTRVLGNKKGVFTROPKSAAFILREYWK IANETRYPHSIAKSQCLENSPFT"
	polyA_signal
	2103. .2108
	polyA_signal
	2126. .2131
ORIGIN	
Query Match	10.3%; Score 174.4; DB 9; Length 2155;
Best Local Similarity	51.6%; Pred. No. 1.4e-35;
Matches	487; Conservative 0; Mismatches 436; Indels 21; Gaps 3;
QY	737 TGGAACCTGAAAAAGACGAGTACACTCTGACATCGGAATCAGAACGATCAGCTGGAGC 796
DB	930 TGGGCGCTGTGTGACTTCTACACCTCCCTGTGGGATCCGACTGTGCTGCACCG 989
QY	797 AGAAGAGCTCTATCTGAACGGGAAACCTCTTTTGAAGGGCTTTGAAAGCAGAG 856
DB	990 AAAGCAGTTCCTCATCAATGGGAAACCTTTCTATTTCACCGGTGTCACAAGCATGAGG 1049
QY	857 AATTCCCGTCTTGGGAGGGGACCTTTTATTCATGTGATGATAAAGACTTCAACCTTC 916
DB	1050 ATGCGACATCCGAGGAGGGCTTGACTGGCCGCTGTGTGAAGAGATTCAACCTGC 1109
QY	917 TGAAGTGAATCAACGGAATCTTTACAGGACCTCTCACTATCCTTACAGTGAAGAGTGC 976
DB	1110 TTCGCTGCTGTGTGCAATGCTTCCGACACGCACTACCCCTACGCGAGGAAGTGC 1169
QY	977 TGATCTTGCAGACAGACTCGGAATCTTGTGATAGACGAGACCCCGACAGTGTGATCA 1036

Db 1170 TGCAGATGTGTGACCCGCTATGGAGTTGTGTGTCATCGATGAGTGTCTCGCGTGGCCTGG 1229  
 QY 1037 CAAGGTACCA--CTACATCCCGAGACTCAGAGATAGCAGAAACAACATAAGAA 10933  
 Db 1230 CACTGCCCGCAGTCTTCAACAACGTCGTCTCGAGAACCATGCGGTGATGAGGAAG 12899  
 QY 1094 TGATCGACAGACACAGAACCATCCAGTGTGATCATGTGAGTGTGGCGAACAACAG 11533  
 Db 1290 TGTGCGCAGGAGCAAGAACCAACCCCGCCTGCTGATGTGTGTGTGTGGCAACGCTTG 13499  
 QY 1154 AGTCCAACCATCCAGACGCGGAGGGTTCTTCAAGCCCTTATGAGACTGCCATGAAA 12133  
 Db 1350 CGTCCCACTAGAACTGCGCGCTACTACTTGAGATGTGTACTACACCAAAACCT 14099  
 QY 1214 TGGATCGAACACGCCCCGTGTGTCATGTGTGAGCATGATGAGCGCACCAAGAGAACAA 12733  
 Db 1410 TGGACCCCTCCCGGCTGTGACCTTGTGACCACTCCAACATATGACG----- 14577  
 QY 1274 GAGACGTGGCGCTGAAGTACTTGCACATGCTGTCGTGTGAACAGAGTACTACGCGTGTACA 13333  
 Db 1458 CAGACAAAGGGGCTCCGATATGTGACGTGATCTGTTGAACAGCTACTACTCTTGTGATC 15177  
 QY 1334 TCTATCAGGGAAGATAGAAGAGGACTTCAAGCTTGAAAAAGACATAGAAGACTCT 13933  
 Db 1518 ACGACTATGGGACACCTGGAGTTGATTAGCGGACGCTTACCAACCACTTGAAGACTGCT 15777  
 QY 1394 ATGCAAGGCACAGAAAGCCCATCTTTGTCCACAGAAATTCGTGGCGACGCGATAGCTGCA 14533  
 Db 1578 ATAAAGCGTATCAGAAAGCCCATTTATCAGAGCGAGTATGAGCGGAAACGATGTTGGGT 16377  
 QY 1454 TCCACTACGATCCACTCAATGTCTCCGAAGAGTACCAAGCAGAGCTCGTTGAAAAA 15133  
 Db 1638 TTACCAAGGACCCGCTCTGATGTTCACTGAAGAGTACAGAGAGTCTGTAGAGCAGT 16977  
 QY 1514 CGATCAGGCTCCTTTTGA AAAA-----AGACTACATCATCGGAACACACGCTGTGGCCT 15677  
 Db 1698 ACCATGTGTTCTTGATCAAAAAACGACGAAAGTACGTGTTGAGAGCTCACTGGAATT 17577  
 QY 1568 TTGCAGATTTTAAGACTCCTCAGAAATGTGAGAAAGCCCATTTCAACCAAGAGGTGTTT 16277  
 Db 1758 TTGCCGATTTTCAATGACTGAACAGTCAACGACGAGAGTGTGGGGAATAAAAAGGGGTCT 18177  
 QY 1628 TCACAAGAGACAGACAAACCAAACTCGTTGCTCATGTACTGAGA 1671  
 Db 1818 TCACTCGGACAGACAAACCAAAAAGTCAAGCCGTTCTTTGGCA 1861

RESULT 7  
 CQ575703 2128 bp DNA linear PAT 02-FEB-2004  
 LOCUS  
 DEFINITION Sequence 3461 from Patent WO0171042.  
 ACCESSION CQ575703  
 VERSION CQ575703.1 GI:41639115  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Drosophila sp.  
 Drosophila sp.  
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 Ephydroidea; Drosophilidae; Drosophila.  
 REFERENCE  
 AUTHORS Venter, J.C., Adams, M., Li, P.W. and Myers, E.W.  
 TITLE Detection kits, such as nucleic acid arrays, for detecting the  
 expression of 10,000 or more Drosophila genes and uses thereof  
 JOURNAL Patent: WO 0171042-A 3461 27-SEP-2001;  
 PE Corporation (NY) (US)  
 FEATURES  
 source location/Qualifiers  
 1. 2128  
 /organism="Drosophila sp."  
 /mol\_type="unassigned DNA"  
 /db\_xref="taxon:7242"

	Best Local Similarity	51.6%	Pred. No. 1.8e-35;	
	Matches	492;	Conservative	0; Mismatches 435; Indels 27; Gaps 3;
QY	734	AGGTGGAACCTTGAAAAAGACGATACACTTGGACATCGAATTCAGAAGCATCAGCTGGG	793	
Db	1104	AGTTGGAATCTTCGACGATACCCTACCGCTCCAGTGGGCATTAGCTTAAGTTGGG	1163	
QY	794	ACGAAAGAGGCTCTATCTGAA CGGAAA CCGTCTTTT TGAAGGCTTTGGAAA GCACG	853	
Db	1164	ACAACGACAGCCTGCTCTGTAATGMAAACCCCTCTATCTGCCGGGATTGGACGCGACG	1223	
QY	854	AGGAATTCCTCCGTTCTGGGCGAGGGCACCTTTATCCATGATGATAAAGACTTCAACC	913	
Db	1224	AGGACTCCGATATCCGGGGAGGGATTGGAATATGCGCTTCTTGCTAGAGATTTTAAAC	1283	
QY	914	TTCTGAAGTGGATCAACGCGCAATCTTTTCAAGGACCTCTCACTATCTTACAGTGAAGAT	973	
Db	1284	TGCTGAAGTGGACTGGAGCCAATGATATCGCACTCTCACTATCTTATTCGMAAGAT	1343	
QY	974	GCGTGGATCTTGCCGACAGACTCGGAATCTTGTAATAGAGAAAGCCCCGACGTTGTA	1033	
Db	1344	CAATGCAGTTTGCCGATCAGCATGGTATTAATGATTATGACGAATGCC-----TGCTG	1397	
QY	1034	TCACAAGGTACCACTACAATCCCAGACTCAGAAAGATACGAAAGACAATTAAGAGAA	1093	
Db	1398	TCATATATAGATATCTTCGAGCCGACGCTACTGGAATACATGTCTCTGCTGGAGCAAC	1457	
QY	1094	TGATCGACAGACACAAGAACCATCCCAAGTGTATCATATGGAAGTGTGGCGAAGCAACAG	1153	
Db	1458	TGATCCACCCGGGACAGAAACCATCCCAAGTGTGTGATGTGCGGTAGCCCAAGAGCCGA	1517	
QY	1154	AGTCCAAACCATCCAGACGGGAGGGTTCTTCAAAGCCCTTATGAGACTGCCAATGAAA	1213	
Db	1518	GATCGAACAAAGCAGGAGGCCCTTAATACTTTGAATTCCTGTAACATATGTAAGAGAAA	1577	
QY	1214	TGGATCGAACACGCCCCGTGTGTATGATGAGCATGATGAGACGACACGAGAGAACAA	1273	
Db	1578	TAGCTCAGGACGACCTCTAAACGCGGCTATAATGCCAATC-----TTCCA	1625	
QY	1274	GAGACGTGGCGCTGAAGTACTTTCGACATCGTCTGTGTAACAGTACTACGCTGTGTA CA	1333	
Db	1626	GTTGCCATTTGGCGCAGTTTCTTGACATCGTGGGTTCAATGCTATACTCTTGATATC	1685	
QY	1334	TCTATCAGGGAAGATAGAAAGACTTCAAGCTCTGGA AAAAGACATAGAAGAGCTCT	1393	
Db	1686	AGAACTCGGCCGCTACTGACATGATACTTAACCTGTGACGATAGAGGCCAGAGTTGGC	1745	
QY	1394	ATCGAAGGACAGAAAAGCCCATCTTTGTCACAGAAATTCGGTGCAGCGGATAGCTGGCA	1453	
Db	1746	GGGATAGGTTGGAAAAGCCTGTCAATTCGAATTTAGTAGCGGCGGACACTATGAGGGCA	1805	
QY	1454	TCCACTACGATCCACCTCAATGTTCTCCGAAAGAGTACCAGCAGAGCTCGT-----	1505	
Db	1806	TGCACCTACTCCCGCTTTATTATTGGTCGAGGAATAACAGTTGAGCTCTTCTTCGCC	1865	
QY	1506	-TGAAAAGACGATCAGGCTCTTTTGAAAAAAGACTACATCATCGGAACACACGCTGGG	1564	
Db	1866	ATTTCAGGCTTTTCGACGAGTTACGCGGAAAGAGATGTTTATTCGAGAGTTGTTTGA	1925	
QY	1565	CCTTTCGACATTTTAAGACTCTCTCAGAAATGTGAGAAAGACCCATTCTCAACCAAGGGTG	1624	
Db	1926	ACTTCGCCGATTTTCGACCGGCGCAGACTATTAACCGCGGTGGCGGCAACAAAAGGGAG	1985	
QY	1625	TTTTTACAGAGACAGACAACCCAAACTCGTTGCTCATGTACTGAGAAAGACTGT	1678	
Db	1986	TCTTTACAAGGAACCGACAAACCCAAAGAAAGTGGCTCAATTCTTAAGGCGCGCGT	2039	



ACCESSION BC014142  
VERSION BC014142.2 GI:40225978  
KEYWORDS MGC.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 2182)  
Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D., Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hejblum,F., Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L., Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S., Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J., Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J., McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W., Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahy,J., Helton,E., Kettman,M., Madan,A., Rodriguez,S., Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y., Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D., Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M., Butlerfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E., Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.  
Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences  
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)  
12477932  
2 (bases 1 to 2182)  
Strausberg,R.  
Direct Submission  
Submitted (10-SEP-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
On Dec 19, 2003 this sequence version replaced gi:15559559.  
Contact: MGC help desk  
Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
Tissue Procurement: ATCC  
cDNA Library Preparation: Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC),  
Gaithersburg, Maryland;  
Web site: <http://www.nisc.nih.gov/>  
Contact: [nisc\\_mgc@nhgri.nih.gov](mailto:nisc_mgc@nhgri.nih.gov)  
Akher,N., Ayale,K., Beckstrom-Sternberg,S.M., Benjamin,B., Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S., Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P., Hansen,N., Ho,S.-L., Karins,E., Kwong,P., Laric,P., Legaspi,R., Maduro,Q.L., Mastello,C., Maskeri,B., Mastrian,S.D., McCloskey,J.C., McDowell,J., Pearson,R., Stancirpop,S., Thomas,P.J., Touchman,J.W., Tsugeon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L., Young,A., Zhang,L.-H. and Green,E.D.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>  
Series: IRAL Plate: 30 Row: 0 Column: 15  
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 4504222.  
location/Qualifiers  
1. 2182  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="MGC:20659 IMAGE:4662011"  
/tissue\_type="Colon, adenocarcinoma"  
/clone\_id="NIH\_MGC\_15"  
/lab\_host="DH10B-R"  
/note="Vector: pOTB7"

gene	
1. 2182	/gene="GUSB" /note="synonym: MPS7" /db_xref="LOCUSID:2990" /db_xref="MIM:253220"
CDS	
2. 1957	/gene="GUSB" /codon_start=1 /product="glucuronidase, beta" /protein_id="AAH14142.1" /db_xref="GI:15559560" /db_xref="MIM:253220"
/translation="MARGSAVAMALGPLMGCALGLQGMLYPQSPSRECKELDGLWSFRADPSDNRRRGFEEQWYRRPLWESGPTVDMPPSSFNDISQWRLRHFGWVWE REVILPERWTQDLRTVRVLRIGSAHSYAIWVNGVDTLEHGGYLPFEADISNVQVQ LPSRLRITIAINNTLPTTLPGRTIQLTDTSKYPKGYFQNTYDFENYAGLQRSV LVTPTTYIDIDITYTTSVEQDSGLVNYQISVKSNNLFKLEVRLLDAENKVVANGTG OGLNGKPFYFPGVAKHEDADIRGKGFDMPLVKDFNLRLGANAFTSHYPAEYV MOMGRYGVIVIDECPGVALPOFPNNVSLHHHQMVEEVRDKHPDAVMMWSVAN EPASHLESAGYLYKVIATKSLDPSRPVTFVSNYAADKGAIFYDVICLNSYSWY HDYGLHLIQLQATQFENYKRYKQPIIOSYGAETIAGFHODPLMFTBEYQKSL EQYHGLDQKRRKYVVELIWNFADFMTEQSPTRVLGNKKGIIFTRQRPKSAFLRE RYWKIANETRYPHSVAKSQCLENSLEFLT"	
ORIGIN	
Query Match	10.3%; Score 174; DB 9; Length 2182;
Best Local Similarity	52.2%; Pred. No. 1.8e-35;
Matches 493; Conservative 0; Mismatches 430; Indels 21; Gaps 4;	
QY	737 TGGAACTTGAAGAAAGACGAGTACACTCTGGACATCGGAATCAGAACGATCAGCTGGGACG 796
DB	939 TGGGGCCTGTGTCTGACTTCTACACACTCCCTGTGGGATCCGCACGTGGCTGCACCA 998
QY	797 AGAAGAGGCTCTATCTGAACGGGAAACCTGCTTTTGAAGGGCTTTGGAAGCAGCAGG 856
DB	999 AGAGCCAGTTCCTCATCAATGGGAAACCTTCTATTCCACGGTGTCAACAACATGAGG 1058
QY	857 AATTCCTGCTGGGGCAGGCACTTTTATTCATTGATGATTAAGACCTTCAACCTTC 916
DB	1059 ATGGGACATCCGAGGGAAGGCTTCGACTGGCCGCTGTGTAAGACCTTCAACCTGC 1118
QY	917 TGAAGTGATCAACGCGAATCTTTCAGGACCTCTCACTATCTTACAGTGAAGAGTGC 976
DB	1119 TTCGCTGGCTGGTGCCAACGCTTTCGTACAGCACCACTACCTATGACAGGAAGTGA 1178
QY	977 TGGATCTTCCGACAGACTCGGAATCCTTGTGATAGACGAAGCCCGCACGTTGTTATCA 1036
DB	1179 TGCAAGTGTGACCGCTATAGGATTTGTGTATCATGATGATGTGCCGCGTGGCCTCG 1238
QY	1037 CAAGGTACCA---CTACAATCCGAGACTCAGAAGATAGACGAAGACAAACATAGAAGAA 1093
DB	1239 CGCTGCCGAGTTCTTCAACAACGTTTCTGTGCATCACCAATGACAGTGAAGAAAG 1298
QY	1094 TGATCGACAGACACAAGAACCAATCCAGTGTGATCATGTGAGTGTGGCGAAGCAACAG 1153
DB	1299 TGCTGCTAGGACACAAGAACCAACCCCGCGTGTGATGTGTGTGGCCAAAGACCTG 1358
QY	1154 AGTCCAACCATTCAGACGCGGAGGTTTCTTCAAGCCCTTATGAGACTGCCAATGAAA 1213
DB	1359 CGTCCACCTAGATCTGTGCTACTACTGAAGATGTGATCGCTCACACCAATCTCT 1418
QY	1214 TGGATCGAACACGCCCCGTTGTGATGATGATGATGACGACCAAGACGAGAGAACA 1273
DB	1419 TGGACCCCTCCCGGCTGTGACCTTGTGAGCAACTTAATATGACG----- 1466
QY	1274 GAGACGTGGCGCTGAAGTACTTGACATCGTGTGTGAACAGGTACTACGGCTGTGATA 1333
DB	1467 CAGACAAGGGGCTCCGTATGTGATGTGATGTGTTGAACAGCTACTACTTGTGTATC 1526
QY	1334 TCTATCAGGGAAGATAGAGAAGACTTCAAGCTCTGGAAGAAAGACATAGAAGACTCT 1393



Db 1527 ACGACTACGGGCGACCTGGAGTTGATTTCAGCTGCGACCTGGCCACCAGTTTGAGAACTGGT 1586  
Qy 1394 ATGCAGGCGACAGAAAGCCATCTTGTTCACAGAAATTCGGTGGCGGCGATAGCTGGCA 1453  
Db 1587 ATAGAAGTATTCAGAAAGCCATTATTTCAGACGAGTATGAGCAGAAACGATTGACAGGGT 1646  
Qy 1454 TCCACTACGATCCACCTCAAAATGTTCTCCGAGAGTACCAAGCAGAGCTGTTGAAAAG- 1512  
Db 1647 TTCACCAAGATCCACCTCTGATGTTTCACTGAGAGTACCAAAAAGTCTGCTAGAGCAGT 1706  
Qy 1513 ACGATCAGGCTCTTTGAAAAA-----AGACTACATCATCGGAACACACAGCTGTGGCCT 1567  
Db 1707 ACCATCTGGTCTGGATCAAAAACGACAGAAATACTGTTGAGAGCTCATTTGGAATT 1766  
Qy 1568 TTGCAGATTTTAAGACTCTTCAGAAATGTGAGAGACCCATTTCTCAACCACAGGGGTGTTT 1627  
Db 1767 TTGCCGATTTTCACTAGTGAACAGTCCACCGACGAGAGTGTGGGGAATAAAAAGGGGATCT 1826  
Qy 1628 TCACAGAGACAGACCAACCAACTGTTGCTCATGTACTGAGA 1671  
Db 1827 TCACTCGGACAGACCAACCAAAAGTGCAGCGTTCCTTTGCGA 1870

RESULT 9  
BT010073 2321 bp mRNA linear INV 12-AUG-2003  
LOCUS Drosophila melanogaster LD10588 full insert cDNA.  
DEFINITION BT010073  
VERSION BT010073.1 GI:33589549  
KEYWORDS FLI\_CDNA.  
SOURCE Drosophila melanogaster (fruit fly)  
ORGANISM Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Ephydroidea; Drosophilidae; Drosophila.  
1 (bases 1 to 2321)

REFERENCE  
AUTHORS Stapleton,M., Brokstein,P., Hong,L., Agbayani,A., Carlson,J.,  
Champe,M., Chavez,C., Dorsett,V., Dresnek,D., Farfan,D., Frise,E.,  
George,R., Gonzalez,M., Guarin,H., Krommiller,B., Li,P., Liao,G.,  
Miranda,A., Mungall,C.J., Munoz,J., Pacleb,J., Paragas,V., Park,S.,  
Patel,S., Phouanavong,S., Wan,K., Yu,C., Lewis,S.E., Rubin,G.M.,  
and Celniker,S.  
TITLE Direct Submission  
JOURNAL Submitted (12-AUG-2003) Berkeley Drosophila Genome Project,  
Lawrence Berkeley National Laboratory, One Cyclotron Road,  
Berkeley, CA 94720, USA  
COMMENT Sequence submitted by:  
Berkeley Drosophila Genome Project  
Lawrence Berkeley National Laboratory  
Berkeley, CA 94720

This clone was sequenced as part of a high-throughput process to  
sequence clones from Drosophila Gene Collection 1 (Rubin et al.,  
Science 2000). The sequence has been subjected to integrity checks  
for sequence accuracy, presence of a polyA tail and contiguity  
within 100 kb in the genome. Thus we believe the sequence to  
reflect accurately this particular cDNA clone. However, there are  
artifacts associated with the generation of cDNA clones that may  
have not been detected in our initial analyses such as internal  
priming, priming from contaminating genomic DNA, retained introns  
due to reverse transcription of unspliced precursor RNAs, and  
reverse transcriptase errors that result in single base changes.  
For further information about this sequence, including its location  
and relationship to other sequences, please visit our web site  
(http://fruitfly.berkeley.edu) or send email to  
cdna@fruitfly.berkeley.edu.

FEATURES  
Source  
1. .2321  
/organism="Drosophila melanogaster"  
/mol\_type="mRNA"  
/strain="y; cn bw sp"  
/db\_xref="taxon:7227"  
1. .2321  
/gene="CG2135"  
gene

CDS  
/db\_xref="FLYBASE:FBgn0039874"  
64..2124  
/gene="CG2135"  
/note="longest ORF"  
/codon\_start=1  
/product="LD10588p"  
/protein\_id="AAQ22542.1"  
/db\_xref="GI:33589550"  
/db\_xref="FLYBASE:FBgn0039874"  
/translation="MKSHMHPVTCLLGSLQVLAYTENDFLVNANRKLDSPLFEY  
DNVRKPTPTKGLYPRDSETRVRSQVLRSDPYDPLQIREKWFMDALRKTR  
EISMVPASVNDITDMLRDHVGTVWERTFFVPRSKMTQRTWLFSSVHSAVW  
INGRNATSHSIGHLPFSEISGLSFGDNRIITWCDNLSNRTIPQGSYKVAIDDG  
TVPIQSYTFDFENYAGIHSVHLVTTPLHISDLEVTQLTKEGLRIDYRWLDAK  
EGLQIQPIQLRVQLRDQGHVAAQINKAVYHGTLLVPNATPWWPYLHSDPYLYNL  
QFELFVANSKEKLESIDQTYRLPVGIRLSWDNDSLLNGKPLYLRGFEHDSDIRG  
KGLDNALILARDENLILKWTGANAYRTSHYPSSESMQFADQHGIMIIDECPVNIDIFE  
POLLENHMSLEQLIHRBRNHPVSVANVANEPRSNKGALKYFEFLVNVREIAHGR  
PLTAIINANSSSCHLAQPLDIVGFNRYSWYQNSGRITMILNLTIEQSWDRFGKP  
VIOFEYGDITMEGMSLPAFIWSEYQVELFSRHFKAPELDRGMFIFGEFVWFADF  
RTAQITIRVGNKKGVFTNRQPKVEVAHILRRYFALAKELDMFGLPBDLTVYSKHI  
HNEL"

ORIGIN

Query Match 10.3%; Score 174; DB 3; Length 2321;  
Best Local Similarity 51.6%; Pred. No. 1.8e-35;  
Matches 492; Conservative 0; Mismatches 435; Indels 27; Gaps 3;

Qy 734 AGGTGAACCTTGAAAAAGACGAGTACACTCTGACATCGAATCGAAGCATGACGTGGG 793  
Db 1100 AGTTGAATCTTCGACAGATACCTACCGCTCCAGTGGCATAGCTTAAGTTGGG 1159  
Qy 794 ACGAAGAGGCTCTATCTGAAACGGAAACCTGTCTTTTGAAGGCTTTGAAACGACG 853  
Db 1160 ACAACGACAGCCTGTCTGTAATGAAAAACCTCTATCTGCGGGAATTGACGCGACG 1219  
Qy 854 AGGATTCCTCCGTTCTGGGCGAGGCACTTTATTCATTGATGATTAAGACTTCAACC 913  
Db 1220 AGGACTCCGATATCCGCGGAAGGATTTGATATGCGCTTCTGCTAGAGATTTAAAC 1279  
Qy 914 TTCTGAAGTGATCAACGCGAATCTTTGAGAGCTCTCACTATCTTACGTAGAAGT 973  
Db 1280 TGCTGAAGTGACTGAGACCAATGATATCGCACTCTCACTATCTTATTCGGAAGT 1339  
Qy 974 GGCTGATCTTGGCGACAGACTCGAATCTTGTGATAGACGAAGCCCGACGTTGTA 1033  
Db 1340 CAATGCAATTTGGCGATCAGCATGTATTATGATTAATGACGAATGCC-----TGCTG 1393  
Qy 1034 TCACAAGGTACCACTACAATCCGAGACTCAGAGATAGCAGAAACAACATAAGAA 1093  
Db 1394 TCAATATAGATATTTTCAGCCGACGACTACTGAGAAATCACATGTCTCGTGAGCAAC 1453  
Qy 1094 TGATGACAGACACAAGAACCATCCAGTGTGATCATGTGAGTGTGGCAAGCAACGAG 1153  
Db 1454 TGATCCACCGGACAGAAACCATCAAGTGTGTGATGATGTGCGTAGCCAAACGAGCCGA 1513  
Qy 1154 AGTCCAACCATCCAGACGCGGAGGTTTCTTCAAGCCCTTTATGAGACTGCGCAATGAAA 1213  
Db 1514 GATGAAACAAGGAGGAGCCCTTAATTAATTGAAATTCCTGTAAACTATATGAGAGAAA 1573  
Qy 1214 TGGATCGAACACGCCCCGTTGTCATGTGAGCATGATGACGCAACGACGAGAGAA 1273  
Db 1574 TAGCTACGAGACGACTTAAACGCGGCTATTAATGCCAACTC-----TTCCA 1621  
Qy 1274 GAGACGTGGCGTGAAGTACTTGCATCGATCGTGTGTGAACAGTACTACGGCTGTACA 1333  
Db 1622 GTTGCATTGTGGCGAGTTTCTGACATCGTGGGTTTCAATCGCTATAACTCTTGGTATC 1681  
Qy 1334 TCTATCAGGAAGATAGAAGAGACTTCAAGCTCTGAAAAAGACATAGAAGAGCTCT 1393  
Db 1682 AGAAGCTGGGCGCTACTGACATGATTAACCTGTTGACGATAGAAGGCCAGAGTTGGC 1741  
Qy 1394 ATGCAAGGCACAGAAAGCCATCTTTGTACAGAAATTGCTGGCGGACGCGATAGCTGCA 1453

Db 1742 GGGATAGGTTGCGAAGAGCTGTGATTCATTTGAGTACGGCGGACACTATGAGGGCA 1801  
Qy 1454 TCCACTAGATCCACCTCAAAATGTTCTCCGAAGATACCAAGAGAGCTCGT----- 1505  
Db 1802 TGCACCTACCTCCCGCTTATTTGGTTCGAGAGATACCAAGTGAAGCTCTTCTCGCC 1861  
Qy 1506 -TGAAAAGACGATCAGGCTCCTTTGAAAAAGACTACATCATCGGAACACACGTTGGG 1564  
Db 1862 ATTCAAGGCTTTCGACGAGTTACGGGAAGAGATGGTTATGGAAGTTGTTGGA 1921  
Qy 1565 CCTTTCAGATTTTAAGACTCTCAGATGTGAGAAGACCCATCTCAACCAAGGGTG 1624  
Db 1922 ACTTCGCGGATTTTCGACCGCGAGACTATTACCCGCTGGCGGCAACAAGGGAG 1981  
Qy 1625 TTTTCAAGAAGACAGAACCAACTCGTTGCTCATGTACTGAGAAGACTGT 1678  
Db 1982 TCTTTACAAGAACCGACCAACCAAGAAGTGGCTCACATTCTTAGCGCGGT 2035

RESULT 10  
AR344264  
LOCUS AR344264 1956 bp DNA linear PAT 17-AUG-2003  
DEFINITION Sequence 1 from patent US 6582692.  
ACCESSION AR344264  
VERSION AR344264.1 GI:33740191  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 1956)  
AUTHORS Podsakoff, G. and Watson, G.  
TITLE Recombinant adeno-associated virus virions for the treatment of  
lysosomal disorders  
JOURNAL Patent: US 6582692-A 1 24-JUN-2003;  
FEATURES  
source 1. 1956  
/organism="Unknown"  
/mol\_type="genomic DNA"

ORIGIN

Query Match 10.2%; Score 172.4; DB 6; Length 1956;  
Best Local Similarity 52.1%; Pred. No. 4.9e-35;  
Matches 492; Conservative 0; Mismatches 431; Indels 21; Gaps 4;  
Qy 737 TGGAACTTGAAGAAAGACGAGTACACTCTGGACATCGGAATCAGAACGATCAGCTGGGACG 796  
Db 938 TGGGGCTGTGCTGACTTCTACACACTCCCTGTGGGATCCGACTGTGCTGTACCA 997  
Qy 797 AGAAGAGGCTCTATCTGAACGGGAACCTGCTTTTGAAGGGCTTTGMAAGCAGAGG 856  
Db 998 AGAGCCAGTTCTCTATCAATGGGAACCTTCTATTTCCACGGGTCAACAAGCATGAGG 1057  
Qy 857 AATTCCCGTTCTGGGGCAGGGCACTTTATCATGTGATGATAAAGACTTCAACCTTC 916  
Db 1058 ATGCGGACATCCGAGGGAGGGCTTGAAGTGGCGCTGCTGTGAAGAGACTTCAACCTGC 1117  
Qy 917 TGAAGTGAATCAACGCAATTTCTTCAAGACCTCTCACTATCCTTACAGTGAAGTGGC 976  
Db 1118 TTCGCTGGCTTGGTCCAAACGCTTCCGTACCAAGCACTACCTATGCAAGAGAGTGA 1177  
Qy 977 TGGATCTTGGCCGACAGACTCGGAATCTTGTGATAGACGAAGCCCGCAGCTGGTATCA 1036  
Db 1178 TGCAGATGTGACCGCTATGGGATGTGTGATCATGATGATGTGCCGCTGGGCTGG 1237  
Qy 1037 CAAGGTACCA--CTACAATCCGAGACTCAGAGATAGCAGAGACAACATAGAAGAA 1093  
Db 1238 CGCTGCGCAGTTCTTCAACAACGTTTCTCTGATCAACACATGCAAGTGAAGAAG 1297  
Qy 1094 TGATCGACAGACAAGAACCATCCAGTGTGATCATGTGAGTGTGGCAAGCAACAG 1153  
Db 1298 TGGTGCGTAGGGAACAAGAACCAACCCCGCGTCTGTGATGTGTGTGTGGCAAGCCTG 1357

Qy 1154 AGTCCAAACCATCCAGACCGGAGGGTTCTTCAAAAGCCCTTTATGAGACTGCCAATGAA 1213  
Db 1358 GTTCCACCTTGAATCTGTGCTACTACTTGAAGATGGTGTGATCGCTCACACCAATCCT 1417  
Qy 1214 TGGATCGAACAGCCCCGTTGTCATGTGTGAGCATGATGAGCCACAGACGAGAACA 1273  
Db 1418 TGAACCCCTCCCGCTGTGACCTTTGTGAGCAACTTAATATGCAAG----- 1465  
Qy 1274 GAGACGTGGCGCTGAAGTACTTGCACATCGTCTGTGTGAACAGGTACTACCGCTGTACA 1333  
Db 1466 CAGACAAAGGGGCTCCGTATGTGATGTGATGTGTTGAACAGCTACTACTTGTGTATC 1525  
Qy 1334 TCTATCAGGGAAGATAGAAGAGACTTCAAGCTCTGAAAAAGACATAGAAGACTCT 1393  
Db 1526 AGACTACGGGCACTGAGTTGATTCACTGCAAGCTGGCCACCCAGTTTGAACACTGGT 1585  
Qy 1394 ATGCAAGGACAGAAAGCCCATCTTGTGCAGAAATTCGGTGGGACGCGATAGCTGCA 1453  
Db 1586 ATAAAGATATCAGAAGCCCATATTATTCAGAGCGAGTATGAGACAGAAACGATTGACGGGT 1645  
Qy 1454 TCCACTACGATCCACCTCAATGTTCTCCGAGAGTACCAACAGAGCTCGTTGAAAAAG- 1512  
Db 1646 TTCAACAGATCCACCTGTGATGTCTACTGTAAGATACCAAGAAAGTCTGTAGACAGT 1705  
Qy 1513 ACGATCAGGCTCTTTTGAAGAA----AGACTACATCATCGGAACACAGCTGTGGCCT 1567  
Db 1706 ACCATCTGGGTCTGATCAAAAACGAGAAATATGTGTGGAAGACTCATTTGGAATT 1765  
Qy 1568 TTGAGATTTTAAGACTCTCAGATGTGAGAAGACCCATCTCAACCAAGGGTGT 1627  
Db 1766 TTGCGGATTTATGACTGAACAGTCAACGAGAGAGTGTGGGGAATAAAGGGGATCT 1825  
Qy 1628 TCACAAGAGACAGAACCAACTCGTTGCTCATGTACTGAGA 1671  
Db 1826 TCACTCGGAGAGACCAACCAAAAGTGACAGCGTTCCTTTGGCA 1869

RESULT 11

AX147652  
LOCUS AX147652 1956 bp DNA linear PAT 08-JUN-2001  
DEFINITION Sequence 1 from Patent WO0136603.  
ACCESSION AX147652  
VERSION AX147652.1 GI:14346708  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE 1 (bases 1 to 1956)  
AUTHORS Podsakoff, G., Watson, G., Couto, L.B. and Yang, B.  
TITLE Recombinant adeno-associated virus virions for the treatment of  
lysosomal disorders  
JOURNAL Patent: WO 0136603-A 1 25-MAY-2001;  
Avigen, Inc. (US); Children's Hospital Medical Center of Northern  
California (US)  
FEATURES  
source 1. 1956  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"  
1. 1956  
/note="unassigned protein product"  
/codon\_start=1  
/protein\_id="CAC41170.1"  
/db\_xref="GI:14346709"

CDS

/translation="MARGSAVWMAALGPLWGCALGLOGMLYPOESPSRECKELDGL  
WSFRADFSNDRRGFEQWYRRPLWESGPTVDMVPSSFNDISQDRLRHFGVWVYE  
REVILPERWTOEDLTRVLRIGSAHSYAIWVNGVDLTLEHGGYLPFEADISNVQVG  
PLPSRLRITAIINNTLTPTTLPPGTIOYLTDRSKYPKGYFVQNTYFDFPNYAGLORSV  
LVTPTTYIDITVTTSVEODBSLVNYOISVGSNLFKLEVRLLDAENKVVANGTGT  
OGQLKVPVSLWMPYLMHERPAYLISLEVQLTAOTS LGPVSDFYTL PVGIRTVATKS  
OFLINGKPFYFHVGNKHEDADIRGQDFWPLVDFNLRLWLGANAFTSHYPVAEEV  
MOMCDRYGIVVIDECPGVGLALPQFPNNVSLHHNMQVMEEVVRDKNPAVVMASVAN

EPASHESAGYVYLNKVIATHTKSLDPSRPVTFVSNNSVYADKGAPYVDVICLNSYYSW  
HDYHLELIQLOLATOENWYKYOKEPIIOSEYGAETIAGFHODPLMFTEEYOKSL  
EQYHGLDQKRRKYVVGELIMNFADPMTQSPTRVLGNKKGIFTRORQPKSAFLIRE  
RYWKIANETRYPHSVAKSQCLENSPFT"

ORIGIN

Query Match	10.2%;	Score 172.4;	DB 6;	Length 1956;
Best Local Similarity	52.1%;	Pred. No. 4.9e-35;		
Matches	492;	Conservative 0;	Mismatches 431;	Indels 21; Gaps 4;
Db	737	TGGAAGCTTGAAGAAAGACGAGTACCTTGACATCGGAATCGAAGATCAGCTGGGACG	796	
Db	938	TGGGGCCTGTGTCTGACTTCTACACACTCCCTGTGGGATCCGCACTGTGGCTGCACCA	997	
QY	797	AGAAGAGGCTCTATCTGAACGGGAAACCTGTCTTTTGAAGGCTTTGAAAGCAGCAGAG	856	
Db	998	AGAGCCAGTTCCTTCATCAATGGGAAACCTTTCTATTTCACGGGTGTCAACAGCATGAGG	1057	
QY	857	AATTCCTCCGTTCTGGGGCAGGGCACTTTATCCATTGATGATTAAGACTTCAACCTTC	916	
Db	1058	ATGCGGACATCCGAGGGAGGGCTTCGACTGGCCGCTGCTGTGTAAGACTTCAACCTGC	1117	
QY	917	TGAAGTGGATCAACCGGAATCTTTCAGGACTCTCCTATCTTACAGTGAAGAGTGGC	976	
Db	1118	TTCCGCTGGCTGTGTGTCACACGCTTTCCGTACACGCACTACCCCTATGCAAGAGAACTGA	1177	
QY	977	TGGATCTTGGCCGACAGACTCCGNAATCCTTGATAGACGAAGCCCGCAGCTGTATCA	1036	
Db	1178	TGCAGATGTGTGACCGCTATGGGATGTGTGTCATGATGATGTCCCGCGTGGCCTGG	1237	
QY	1037	CAAGGTACCA---CTACAATCCCGAGACTCGAAGATAGCAGAAAGACAACATAAGAGAA	1093	
Db	1238	CGCTGCCGCACTTCTTCAACAACGTTTCTCTGCATCACCATGCAAGTGATGGAAGAA	1297	
QY	1094	TGATCGACACACAAAGAACCATCCAGTGTGATCATGTGAGTGTGGCGAAACGAACCA	1153	
Db	1298	TGGTGCCTAGGGACAAAGAACCAACCCCGCGTGTGATGTGTGTGTGTGTGTGTGTGT	1357	
QY	1154	AGTCCCAACCATCCAGACGCGGAGGGTTCTTCAAAAGCCCTTTATGAGACTGCCAATGAA	1213	
Db	1358	CGTCCCACTTGAATCTGCTGCTACTACTTGAAGATGTGATCGCTCACACCAATCTCT	1417	
QY	1214	TGGATCGAACACGCCCCGCTGTGATGTGATGATGATGATGATGATGATGATGATGAT	1273	
Db	1418	TGGACCCCTCCCGGCTGTGACCTTGTGTGACCACTTAACTATGACAG-----	1465	
QY	1274	GAGACGTGGCGCTGAAGTACTTGCACATCGTCTGTGTGAAACAGGTACTAGGCTGTACA	1333	
Db	1466	CAGACAAAGGGGCTCCGATGTGATGTGATGTGATGTGATGTGATGTGATGTGATGTG	1525	
QY	1334	TCTATCAGGGGAAGATAGAAAGACTTCAAGCTCTGGAAGAAAGACATAGAAAGCTCT	1393	
Db	1526	ACGACTACGGGACACCTGGAGTGTGATTCAGCTGCAGCTGGCCACCCAGTTTGAAGACTG	1585	
QY	1394	ATGCAAGGCAAGAAAGCCCATCTTGTGTCACGAATTGGGTGGGACGGGATAGCTGCA	1453	
Db	1586	ATAAGAAATATCAGAAAGCCCATTAATTCAGAGCGAGATGAGCAGAAAGATTCAGGG	1645	
QY	1454	TCCACTACCATCCACCTCAAAATGTTCTCCGAAAGAGTACCAAGCAGAGCTCGTTGAAAG	1512	
Db	1646	TTCACCAAGATTCACCTCTGATGTTCATCTGAAAGATACGAGAAAGTCTGCTAGAGCAG	1705	
QY	1513	ACGATCAGGCTCTTTGAAAAA---AGACTACATCATCGGAACACACAGTGTGGGCT	1567	
Db	1706	ACCATCTGGGCTTGATCAAAAAACGAGAAATATGTGTGAGAGCTCATTTGGAAT	1765	
QY	1568	TTGCAGATTTTAAGACTCTCTCAGATGTGAGAGAACCACTTCTCAACCAAGAGGTGTT	1627	
Db	1766	TTGCCGATTTTCACTGTAACATGACCCGACGAGAGTGTGGGAAATAAAAAGGGATCT	1825	
QY	1628	TCACAAGAGACAGACAACCAAACTCGTTGCTCATGTACTGAGA	1671	
Db	1826	TCACTCGGACAGACCAACCAAAAGTGCAGCGTCTTTTGGCA	1869	

RESULT 12

HUMGLCB

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

COMMENT

FEATURES

source

gene

mRNA

CDS

HUMGLCB 2191 bp mRNA linear PRI 08-NOV-1994

Human beta-glucuronidase mRNA, complete cds.

M15182

M15182.1 GI:183232

alternative splicing; beta-D-glucuronoside glucuronosohydrolase;

beta-glucuronidase; glucuronidase; glucuronohydrolase; hydrolase.

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

Oshima,A., Kyle,J.W., Miller,R.D., Hoffmann,J.W., Powell,P.P.,

Grubb,J.H., Sly,W.S., Tropak,M., Guise,K.S. and Gravel,R.A.

Cloning, sequencing, and expression of cDNA for human

beta-glucuronidase

Proc. Natl. Acad. Sci. U.S.A. 84 (3), 685-689 (1987)

87118233

3468507

Original source text: Human fibroblasts cDNA clone pHUGF and

placenta cDNA clones pHUG13 and pHUGP15.

Draft entry and copy of computer-readable sequence of [1] kindly

provided by R.D.Miller, 08-MAY-1987.

Four potential N-linked glycosylation sites are located at bases

543-551, 840-848, 1284-1292 and 1917-1925. Potential poly-A

signals were found 25 and 48 bp upstream of the poly-A site at

nucleotides 2139-2144 and 2162-2167 respectively.

Two types of mRNA exit due to alternative splicing. The short form,

lacking bases 939-1091 below, encodes a protein lacking

beta-glucuronidase activity.

Location/Qualifiers

1. .2191

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/map="7q22"

1. .2191

/gene="GUSB"

<1. .2191

/gene="GUSB"

/product="glucuronidase-b mRNA"

27. .1982

/gene="GUSB"

/note="beta-glucuronidase precursor (EC 3.2.1.31)"

/codon\_start=1

/protein\_id="AA52561.1"

/db\_xref="GI:183233"

/db\_xref="GDB:G00-120-025"

/translation="MARGSAVAMALALGPLWGCALGLOGGMLYPPESPSPRECKELDGL  
WSFRADFSNRRRGEFEQWYRPVSSPTVDMPVSSFNDSODRLRHFGVWWE  
REVILPERWTQDLKRVLRIGSAASYAIWVWNGVDLLEHGGYLPREADISNLVYG  
PLPSRLRITIAINNTLPTTLPPFTIQLTDTSKYPKGYFVONTYDFENYAGLORSY  
LVTPTTYIDDIYVTSVEQDSGLVNYOISVKSNLKLEVRLLDAENKVVANGTGT  
OGQLRVPGVSLWVPLMHERPAYIYSLVOLTAQTSIGPVSDFYTLPGVIRTVATKS  
OPLNGKPFYFHGVNKHEDADIRKGFDMPLVADPNLRLMLGANAARTSHYPABEV  
MOMCDRYGIVIDECPGVGLALPQFNNSLIHHMQMBEVRDKNHAPVMMVSVAN  
EPASHLESAGYVYLNKVIATHTKSLDPSRPVTFVSNNSVYADKGAPYVDVICLNSYYSW  
HDYHLELIQLOLATOENWYKYOKEPIIOSEYGAETIAGFHODPLMFTEEYOKSL  
EQYHGLDQKRRKYVVGELIMNFADPMTQSPTRVLGNKKGIFTRORQPKSAFLIRE  
RYWKIANETRYPHSVAKSQCLENSPFT"

27. .92

/gene="GUSB"

/note="beta-glucuronidase signal peptide"

93. .1979

/gene="GUSB"

/product="beta-glucuronidase mature peptide"

mat\_peptide

sig\_peptide

33 bp upstream of SmaI site.

ORIGIN

Query Match

Best Local Similarity

10.2%;

Score 172.4;

DB 9;

Length 2191;

52.1%;

Pred. No. 4.9e-35;

Matches 492; Conservative 0; Mismatches 431; Indels 21; Gaps 4;			
QY	737	TGGAACCTGAAAAAGACGAGTACCTCTGGACATCGGAATCAGAAAGATCAGCTGGGACG	796
Db	964	TGGGGCTGTGTCTGACTTCTACACACTCCCTGTGGGGATCCGCACCTGTGGCTGCACCA	1023
QY	797	AGAAGAGGCTCTATCTGAACGGGAACTGTCTTTTGAAGGGCTTTGGAAAGACGAGG	856
Db	1024	AGAGCCAGTTCTTCATCAATGGGAACTTTCTATTTCCACGGTGTCAACAGCATGAGG	1083
QY	857	AATTCCCCGTTCTGGGGCAGGCGACCTTTTATCCATGTATGATAAAGACTTCAACTTC	916
Db	1084	ATGCGGACATCCGAGGGAGGGCTTTCGACTGGCCGCTGTGTGAAGGACTTCAACTGC	1143
QY	917	TGAAGTGAATCAACGCGAATCTTTTCAGGACCTCTCACTATCTTACAGTGAAGTGGC	976
Db	1144	TTCCGTGGCTTGTGTGCCAACGCTTTCCTGTACAGCCACTACCCCTATGACAGAGATGA	1203
QY	977	TGGATCTTGGCCGACAGACTCGGAATCCTTGTGATAGACGAAGCCCCGACGTTGGTATCA	1036
Db	1204	TGCAGATGTGTGACCGCTATGGGATGTGTGTCATCATGATGATGTCCGGCGTGGCCTGG	1263
QY	1037	CAAGGTACCA--CTACAATCCCGAGACTCAGAAAGATAGCAGAAACAACATAAGAA	1093
Db	1264	CGCTGCCGAGTTCTTCAACACGTTTCTCTGATCAACACATGACAGGTGATGGAAG	1323
QY	1094	TGATCGACAGACACAAGAACCACCTCAGTGTGATCATGTGAGTGTGGCGAAGCAAG	1153
Db	1324	TGTTGCGTAGGGAACAAGAACCCCGCGGTGTGATGTGTGTGTGGCCACGACCTG	1383
QY	1154	AGTCCAACCATCCAGACGCGAGGGTTTCTTCAAGCCCTTTATGAGACTGCCAATGAAA	1213
Db	1384	CGTCCACCTAGATCTGTGTGCTACTACTTGAAGATGTGATCGCTCACACCAATCCT	1443
QY	1214	TGGATCGAACACAGCCCGCTGTGTGATGAGCATGATGACGCAACGACGAGAGAA	1273
Db	1444	TGGACCCCTCCCGCCTGTGACCTTGTGAGCACTTAATACTATGCAG-----	1491
QY	1274	GAGACGTGGCGCTGAAGTACTTGCATCTGTGTGAACAGGTACTACGGCTGGTACA	1333
Db	1492	CAGACAAAGGGGCTCCGTATGTGATGTGATCTGTTGAACAGCTACTACTCTGTATC	1551
QY	1334	TCTATCAGGAAGATAGAAAGAACTTCAAGCTCTGAAAAAAGACATAGAAGCTCT	1393
Db	1552	ACGACTACGGGCACTGGAGTTGATTCAGCTGCACTGGCCACCGATTGAGAACTGGT	1611
QY	1394	ATGCAAGGACACAGAAAGCCCATCTTGTACAGAAATTCGTGCGGACCGGATAGTGGCA	1453
Db	1612	ATAAGAAGTATCAGAAAGCCATTAATTCAGAGCGATATGAGCAGAAACGATTGACGGGT	1671
QY	1454	TCCACTACGATCCACCTCAATGTTCTCCGAGAGTACCAAGCAGAGCTCGTTGAAAAG-	1512
Db	1672	TTCAACGAGATCCACCTCTGATGTTCACTGAAGAGTACCAAGAAAGTCTGCTAGACAGT	1731
QY	1513	ACGATCAGGCTCCTTTTGA AAAA-----AGACTACATCATCGGAAACACACGTGGGCT	1567
Db	1732	ACCATCTGGGTCTGGATCAAAAAGCCAGAAATATGTGTTGGAGAGCTCATTTGGAATT	1791
QY	1568	TTGCAGATTTTAAAGACTCTCAAGATGTGAGAAAGCCATTCTCAACCAAGGGTCTT	1627
Db	1792	TTGCCGATTTTCATGACTGAACAGTCAACCGACGAGTGTCTGGGAAATAAAAAGGGATCT	1851
QY	1628	TCAACAAGAGACAGAACCCAACTCGTTGCTCATGTACTGAGA 1671	
Db	1852	TCACTCGGACAGACAAACCAAAAGTGACGCTCTTTTGGCA 1895	

RESULT 13  
RNGLCB 2391 bp mRNA linear ROD 10-FEB-1999  
LOCUS Rat mRNA for beta-glucuronidase.  
DEFINITION Y00717  
ACCESSION Y00717.1 GI:56270  
VERSION

KEYWORDS	data-glucuronidase.		
SOURCE	Rattus norvegicus (Norway rat)		
ORGANISM	Rattus norvegicus		
REFERENCE	Powell, P.P., Kyle, J.W., Miller, R.D., Pantano, J., Grubb, J.H. and Sly, W.S.		
AUTHORS	Rat liver beta-glucuronidase. cDNA cloning, sequence comparisons and expression of a chimeric protein in COS cells		
JOURNAL	Biochem. J. 250 (2), 547-555 (1988)		
MEDLINE	88183378		
PUBMED	3355537		
REFERENCE	2 (bases 1 to 2391)		
AUTHORS	Miller, R.D.		
TITLE	Direct Submission		
JOURNAL	Submitted (07-JAN-1988) Miller R.D., E.A. Doisy Dept. of Biochemistry, St. Louis University School of Medicine, 1402 S. Grand Blvd., St. Louis, MO, 63104		
FEATURES	location/Qualifiers		
source	1..2391		
	/organism="Rattus norvegicus"		
	/mol_type="mRNA"		
	/strain="Sprague-Dawley"		
	/db_xref="taxon:10116"		
	/clone="pRLG5"		
	/tissue_type="liver"		
	<1..1909		
	/codon_start=2		
	/product="beta-glucuronidase precursor"		
	/protein_id="CAA68705.1"		
	/db_xref="GI:758260"		
	/db_xref="GOA:P06760"		
	/db_xref="Swiss-Prot:P06760"		
	/translation="ELLCSCALALQGMLEFPKETPSRELKVLDDGLMSFRADYSNNRLQ GFEKQWYRQPLRESGPTLMDPVPSPNDITQEAELRNFIGMWYREBAVLPRWTQDT DRRVLRINSAHYAVVWNGIHVEHEGHLPEADITKLVOGSPVLTFRVTIAINN TLTPYTLPGTIVKTDPSMYPKGFVQDISDFENYAGLHRSVLYTTPTYIDIT VTTDVRDVLGLVNYWISVQSDHFDLEVRLLDEDEKIVARGTGNELQKVPRAHLMWP YLMHHPAYLISLEVMTTPESVSDPYTLPGIRVAVTKSGFLNGKPEYFQGVNKH EDSDIRGRGFDMPLIKDFNLRLWLGANSFRSHYSESVLQLCDRYGIVIDECPG VGIVLPQSGFNSLRHLLEVMDELVRDKNHPAVVMSVANEPVSLKPAGYVFKTLI AHTKALDPTRPVTFVSNTRYDADGAPYVDIVCVNSYLSWYHDYGLLEVTLQQLTSQF ENWYMYQKPIQSEYGADAVSGLHEDPRMFSEETALLENHYLLIDEKREYVIG ELIWNFADFMNTQSDPLRVTKNGKGIFTQRNPKMAAFILRERYWRIANETRGYSVPR TQCMGSRPPTF"		
CDS	<1..28		
	29..1906		
	/product="beta-glucuronidase"		
	2391		
ORIGIN	polya_site		
Query Match	9.9%; Score 166.8; DB 10; Length 2391;		
Best Local Similarity	51.3%; Pred. No. 1.6e-33;		
Matches 477; Conservative	0; Mismatches 432; Indels 21; Gaps 3;		
QY	751	GACGAGTACACTGTGACATCGGAATCAGACGATCAGCTGGGACGAGAGAGGCTCTAT	810
Db	902	GACTTCTATACCTCCCTGTGGGATTGCAACAGTGGCTGTCACAAAGACGAA GTTCCCTC	961
QY	811	CTGAACGGGAAACCTGTCTTTTGAAGGCTTTGAAAGCAGAGGAATTCCTCCGTTCTG	870
Db	962	ATAAATGGGAAGCCTTTCTACTTCCAAGGCGTCAACAAGCATGAGGATTCAGATATCCGA	1021
QY	871	GGGACGGGACACCTTTATCCATTGATGATTAAGAGCTTCAACCTTCTGAAGTGAATCAAC	930
Db	1022	GGGAGAGGCTTGCATGGCCTCTGCTGATTAAGGATTTC AACCTCCTCGGTGGCTCGGG	1081
QY	931	GCGAATCTTTTCAGACCTCTCACTATCTTACAGTGAAGAGTGGCTGATCTTGGCCGAC	990
Db	1082	GCAAAATTCCTTTCGTACCAAGCACTATCCCTACTCGAGAGGATTAATTCAAGCTCTGTGAC	1141



QY	991	AGACTCGGAATCCTTGTGATAGACGAAGCCCCCGACGTTGGTATCACAAGGTACCA---C	1047
Db	1142	CGATATGGAATTGTGGTCATCGATGAGTGTCCGGGTGTGGGCATCGTGTGCCCAAGT	1201
QY	1048	TACAATCCCCGACACTCAGAAAGTAGCAGAAACCAACATAAGAAAGATGATCGACAGACAC	1107
Db	1202	TTTGGCAACGTGTCTCTTCGGCACCACTAGAGGTGATGACGAGCTGGTGGCAGGGAC	1261
QY	1108	AAGAACCATCCCAAGTGTGATCATGTGAGTGTGGCGAACGAAACAGAGTCCCAACCATCCA	1167
Db	1262	AAAAATCACCCCTGCGGTGCTGATGTGCTGTGGCCAATGAGCCTGTCTCTTCTGAAA	1321
QY	1168	GACGCGAGGGTTTCTTCAAGCCCTTATAGACTGCCAATGAATGATCGAACAACGC	1227
Db	1322	CCTGCCGATATTACTTCAAGACGCTGATCGCCACACCAAGCCCTGACCCCAACCGT	1381
QY	1228	CCCGTGTGCATGGTGAAGCATGATGGAACGACCAAGAGAAACAAGAGCGTGGCGCTG	1287
Db	1382	CCTGTGACCTTGTGAGCAATA-----CCAGATATGACGACAGACTGGGGGCC	1429
QY	1288	AAGTACTTGCACATCGTCTGTGTGAACAGGTACTACGCGTGTATCATCTATCAGGAAG	1347
Db	1430	CGTACGTGACGTGATTTGTGTGAACAGTTACTTATCTGTGTATCATGACTACGGGCAT	1489
QY	1348	ATAGAAGAAGGACTTCAAGCTCTGGAAGAAAGACATAGAAGAGCTCTATGCAAGGCACAGA	1407
Db	1490	CTGAGAGTGATTCAGCTGCAGCTGACTAGCCAGTTTGAAACTGTGTATTAAGATGTACCAG	1549
QY	1408	AAGCCCATCTTTGTCAACGAATTGSGTCCGACGCGATAGCTGGCATCCACTACGATCCA	1467
Db	1550	AAGCCAATTATCCAGAGCGAGTATGAGACGACGCGCTCTCGGGCTTCATGAGATCCA	1609
QY	1468	CTTCAAAATGTTCTCCGAAGAGTACCAAGCAGAGCTCGTTGAAAAAGCGATCAGGCTCCTT	1527
Db	1610	CTTCGCATGTTCACTGAGAGTACACACAGACTCTCTCTGGAGAATTATCATTTGATTCCTG	1669
QY	1528	TTGAA-----AAAAGACTACATCATCGGAACACACGTTGGGCTTTGACAGATTTTAAG	1581
Db	1670	GATGAGAAAAGAAAGATATGTATCATCGAGAGCTCATCTGGAATTTTGCTGACTTCATG	1729
QY	1582	ACTCCTCAGAAATGTGAGAAGACCACTTCTCAACCACAAGGGGTGTTTTCACAAGAGACAGA	1641
Db	1730	ACGAACCAAGTACCACTGAGAAGTAAACAGAAAACAAGAAAGGGGATCTTCACTCGACAGAGA	1789
QY	1642	CAACCCAAACTCGTTGCTCATGTACTGAGA	1671
Db	1790	AAACCCAAAGATGGCAGCCTTCAATTTTGCGA	1819

RESULT 14				
AX827491				
LOCUS				
DEFINITION	AX827491	2472 bp	DNA	linear
ACCESSION	Sequence 225 from Patent EP1344834.			PAT 12-DEC-2003
VERSION	AX827491			
KEYWORDS	AX827491.1	GI:39837679		
SOURCE				
ORGANISM	Rattus norvegicus (Norway rat)			
	Rattus norvegicus			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;			
	Rattus.			
REFERENCE	1			
AUTHORS	Boess, F., Suter-Dick, L. and Wolf, D.			
TITLE	Methods for the toxicity prediction of a compound			
JOURNAL	Patent: EP 1344834-A 225 17-SEP-2003;			
	F. HOFFMANN-LA ROCHE AG (CH)			
FEATURES	Location/Qualifiers			
Source	1..2472			

```

/organism="Rattus norvegicus"
/mol_type="unassigned DNA"
/db_xref="taxon:10116"

```

Query Match	9.9%	Score 166.8;	DB 6;	Length 2472;
Best Local Similarity	51.3%;	Pred. No. 1.6e-33;		
Matches 477; Conservative	0;	Mismatches 432;	Indels 21;	Gaps 3;

QY	751	GACGAGTACACTCTGGACATCGGAATCAGAACGATCAGCTGGGACGAGAACGAGGCTCTAT	810
Db	946	GACTTCTATACCTCCCTCGTCGGATTGCAACAGTGGCTGTCACTAAAGCAAGTTCCCTC	1005
QY	811	CTGAACGGGAAACCTGTCTTTTGAAGGCTTTGGAAAGCACGAGGAATTCCTCCGTTCTG	870
Db	1006	ATAAATGGGAGCCCTTTCTACTTCCAGGCGTCAACAGCATGAGATTAGATATCCGA	1065
QY	871	GGGCGAGGCACTTTTATCCATTGATGATATAAAGACTTCAACCTTCTGAAGTGATCAAC	930
Db	1066	GGGAGAGGCTTCGACTGGCCTCTGCTGATTAAGGATTTCAACTCCTCCGTTGGCTCGGG	1125
QY	931	GCGAATTCCTTCAGGACCTCTCACTATCCTTACAGTGAAGAGTGGCTGATCTTGCCGAC	990
Db	1126	GCAATTCCTTTGCTACCGACCACTATCCCTACTCGGAGAGGTACTTCAGCTCTGTGAC	1185
QY	991	AGACTCGGAATCCTTGTGATAGACGAAGCCCCGACGCTGGTATCACAGGTACCA--C	1047
Db	1186	CGATATGGAATTGTGTGATCGATGAGTGTCCCGGTGTGGGCATCGTGTGCCACAGAGT	1245
QY	1048	TACAATCCCGAGACTCAGAAATAGCAGAAACAACATAGAAGATGATCGACAGACAC	1107
Db	1246	TTTGGCAACGTGTCTCTTGCGCACCACTAGAGGTATGACGAGCTGGTCCGACGAGAC	1305
QY	1108	AAGAACCATCCCACTGATGATCATGTGAGTGTGGGAAAGCAACGAGTCCAACCATCCA	1167
Db	1306	AAAAATCACCTGCGGTGATGTGTGTGTGGCCAATGAGCTGTCTCTTCTGTAAA	1365
QY	1168	GACGCGAGGGTTTCTTCAAGCCCTTATAGAGCTGCCAATGAATGATCGAACACGC	1227
Db	1366	CCTGCCGATATTACTTCAAGACGCTGATGCGCCACACCAAGCCCTGACCCACCCGT	1425
QY	1228	CCCCTGTGATGCTGAGCATGATGACGCAACGACGAGAAACAAGAGACGTGGCGCTG	1287
Db	1426	CCTGTGACCTTTGTGAGCAATA-----CCAGATATGACGACATGGGGGCC	1473
QY	1288	AAGTACTTCGACATCGTCTGTGTGAACAGGTACTACGGCTGTATCATTCAGGGAAAG	1347
Db	1474	CCGTACGTGACGTGATTTGTGTGAACAGTTACTTATCCTGTATCATGACTACGGGCAT	1533
QY	1348	ATAGAAGAGGACTTCAGCTCTGAAAAAACAATAGAAGCTCTATGCAAGGCACAGA	1407
Db	1534	CTGAGGTGATTCAGCTGACGCTGACTAGCCAGTTTGAAACTGTATTAAGTGTACCAG	1593
QY	1408	AAGCCCATCTTTGTACACAGATTGCTGCGGACGCGATAGCTGGCATCCACTACGATCCA	1467
Db	1594	AAGCCAAATTATCCAGAGCGAGTATGAGACAGACGCCGCTCTGGGGCTTCATGAGGATCCA	1653
QY	1468	CCTCAATGTTCTCCGAAGAGTACCAAGCAGAGCTCGTTGAAAAAGACGATCAGGCTCCTT	1527
Db	1654	CCTCGCATGTTCAGTAGAGAGTACCAGACAGCTCTCCTGAGAAATTATCATTTGATCCCTG	1713
QY	1528	TTGAA-----AAAAGACTACATCATCGGAACACACGCTGTGGGCTTTGCAGATTTTAAG	1581
Db	1714	GATGAGAAAAGAAAAGATATGTTCATCGGAGAGCTCATCTGGAATTTTGTGCTGACTTCATG	1773
QY	1582	ACTTCCTAGATGTGAGAGACCCATTCTCAACACCAAGGGGTGTTTTTCAACAAGACAGAGA	1641
Db	1774	ACGAAACCAAGTCAACCACTGAGAGTAAACAGAAACAAGAGGGGATCTTCACTCGACAGAGA	1833
QY	1642	CAACCCAAACTCGTTGCTCATGTACTGAGA	1671
Db	1834	AACTCCAAAGATGCAAGCCTTCATTTTGGCA	1863

RESULT 15			
RATGLCB			
LOCUS	RATGLCB	2472 bp	mRNA
DEFINITION	Rat beta-glucuronidase	mRNA, complete cds.	linear
			ROD 27-APR-1993



ACCESSION M13962  
VERSION M13962.1 GI:204329  
KEYWORDS beta-glucuronidase.  
SOURCE Rattus norvegicus (Norway rat)  
ORGANISM Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
REFERENCE 1 (bases 1 to 2472)  
AUTHORS Nishimura,Y., Rosenfeld,M.G., Kreibich,G., Gubler,U.,  
Sabatini,D.D., Adesnik,M. and Andy,R.  
TITLE Nucleotide sequence of rat preputial gland beta-glucuronidase cDNA  
and in vitro insertion of its encoded polypeptide into microsomal  
membranes  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 83 (19), 7292-7296 (1986)  
MEDLINE 87016933  
PUBMED 3463967  
COMMENT Original source text: Rat (Wistar, female) preputial gland, cDNA to  
mRNA, clones p-beta-G[2,5,5'].  
Draft entry and clean copy sequence for [1] kindly provided by  
Y.Nishimura, 19-DEC-1986.  
Beta-glucuronidase undergoes a posttranscriptional proteolytic  
cleavage near its carboxyl-terminal end, which reduces its size by  
approximately 3 kd. The site of this cleavage has as yet not been  
determined.  
FEATURES  
source Location/Qualifiers  
1..2472  
/organism="Rattus norvegicus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10116"  
<1..2472  
/note="g1cb mRNA (alt.)"  
<1..2437  
/note="g1cb mRNA (alt.)"  
7..1953  
/note="beta-glucuronidase precursor"  
/codon\_start=1  
/protein\_id="AAA1228.1"  
/db\_xref="GI:204330"  
/translation="MSRRSVCWFLGQLLCSAVALQGMFLPKETPSRELKLDGL  
WSFRADYNNRLQGFQKQWYRQPLRESGPTLDMVPSSFNDITQBAELNFIQWVWE  
REAVLPQRTQDTRVVLRLNSAHYAVVWVNGIHVVEHGGHLEADITKLVSOG  
PLTFRTVTAIINTLPTPLTPTIVYKTDPSPKGYFVODISDFENYAGLHRSYV  
LYTPTTYIIDITVTVDVDRDGLVNWISVQSDHFOLEVRLLDEDKIVARGTENE  
GOLKVPRAHLMPYLMHEBAYLYSLEVTMTTPESVSDFTLPGVIRTVAVTSKFLI  
NGKPYFQGVNKHEDSDIRGFGDWPLIKDFNLRLWLGANSFRISHYPYSEEVQLC  
DRYGVVIDECPGVGYLVPQSGFVNSLRHLEVMDELVRDKNHPAVVMMSVANEPS  
SLKPYGYFKTLIAHTKALDPTRPVTFVSNTRYDADNGAPYVDICVNSYLSWHDYG  
HLEVIQLOLTSQFENWYKMYKPIIQSEYGADAVSGHEDPRMFSEYQITALLENH  
LILDERKREYVIGELIWNFADFMTNOSPRLVTGNKKGIFTQRNPKMAAFILREYWR  
IANETRGYSVPRTCQMSRPTF"  
7..72  
sig\_peptide  
/note="beta-glucuronidase signal peptide"  
73..1950  
mat\_peptide  
/product="beta-glucuronidase"  
ORIGIN 48 bp upstream to PvuII site.  
Query Match 9.9%; Score 166.8; DB 10; Length 2472;  
Best Local Similarity 51.3%; Pred. No. 1.6e-33;  
Matches 477; Conservative 0; Mismatches 432; Indels 21; Gaps 3;  
QY 751 GACGAGTACACTCTGACATCGAATCAGACGATCAGCTGGACGAGAAGGCTCTAT 810  
DB 946 GACTTCTATACCTCCCTGTCGGGATTCGAAACAGTGGCTGTCACAAAGAGCAAGTTCTC 1005  
QY 811 CTGAACGGGAACTGTCTTTTGAAGGCTTTGAAAGACGAGGAATCCCGTTCG 870  
DB 1006 ATAAATGGAGCCTTCTACTTCCAAAGCGCTCAACAAGCATGAGATTCAATATCCGA 1065  
QY 871 GGGCAGGCGACCTTTTATCCATGATGATAAAGACTTCAACCTTCTGAAGTGATCAAC 930  
DB 1066 GGGAGAGGCTTCACTGCGCTGTGATAAAGATTTCACCTCCTCGTTGGCTCGG 1125

QY 931 GCGAATCTTTCAAGACCTCTCACTATCTTACAGTGAAGAGTGGCTGATCTTGCCGAC 990  
DB 1126 GCAAAATCTTTCTGTACAGACCACTATCCCTACTCGGAGAGGTACTTACGCTCTGTGAC 1185  
QY 991 AGACTCGGAATCCTGTGTATAGACGAAGCCCGCACGTTGTATCACAAGGTACA---C 1047  
DB 1186 CGATATGGAATGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1245  
QY 1048 TACAATCCCGAGACTCAGAAGATAGCAGAACAATGAAGATGATGATGATGATGATGATGAT 1107  
DB 1246 TTGGCAACGTGTCTCTTCCGACCACTAGAGGTGATGACGAGCTGTGCGCAGGAGAC 1305  
QY 1108 AAGAACCATCCAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1167  
DB 1306 AAAAATCACCTGCGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1365  
QY 1168 GACGCGAGGTTTCTTCAAAAGCCCTTATGAGACTGCCAATGAATGATGATGATGATGATGAT 1227  
DB 1366 CTGCGCGATATTACTTCAAGACGCTGATGCGCCACACCAAGCCCTGAGCCCAACCGT 1425  
QY 1228 CCGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1287  
DB 1426 CCGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1473  
QY 1288 AAGTACTTGCATCTGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1347  
DB 1474 CCGTACGTGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1533  
QY 1348 ATAGAAGAAGACTTCAAGCTTGAAGAAAGACATAGAAGCTATGACAGGACACAGA 1407  
DB 1534 CTGAGAGTGTATCAGCTGACGTGACTAGCCAGTTTGAAGCTGTATGAAGTGTACAG 1593  
QY 1408 AAGCCATCTTGTACAGAAATTCGGTGGAGCGGATAGCTGCGATCCACTACGATCCA 1467  
DB 1594 AAGCCAATTATCAGAGCGAGTATGAGACAGACGCGCTCGGGGCTTCATGAGATCCA 1653  
QY 1468 CCTCAATGTCTCCGAAGATACCAAGCAGAGCTCGTTGAAGAAAGACGATCAGGCTCCTT 1527  
DB 1654 CCGCATGTTCAGTGAAGATACCAAGCAGAGCTCCTCGAGAAATTATCATTTGATCCTG 1713  
QY 1528 TTGAA-----AAAAGCTATCATCATCGGAACACACAGCTGTGGCCTTGACAGTTTAAG 1581  
DB 1714 GATGAGAAACGAAAGAAATATGTATCGGAGAGCTCATCTGAAATTTGCTGACTTCATG 1773  
QY 1582 ACTCCTCAGATGTGAGAAGACCACTTCTCAACCAAGGGTCTTTCACAGAGACAGA 1641  
DB 1774 ACGAACCAGTACCACTGAGAGTAACAGGAACAAGAGGGATCTTCACTGACAGAGA 1833  
QY 1642 CAACCCAACTCGTTGCTCATGTACTGAGA 1671  
DB 1834 AACCCAAAGATGCGACCTTCATTTTGGCA 1863

Search completed: January 24, 2005, 05:57:39  
Job time : 7300 secs

This Page Blank (uspto)

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 24, 2005, 07:30:35 ; Search time 25 Seconds  
(without alignments)  
2166.802 Million cell updates/sec

Title: US-09-936-759-6  
Perfect score: 3001  
Sequence: 1 MVRPQRNKKRFFILLINGVWN.....TRDRQPKLVAVHLRLMSEV 563

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 segs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_79: \*  
1: pir1: \*  
2: pir2: \*  
3: pir3: \*  
4: pir4: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2999	99.9	563	2 A72300	beta-glucuronidase
2	1011	33.7	570	2 C90485	beta-glucuronidase
3	899.5	30.0	648	2 A32576	beta-glucuronidase
4	894.5	29.8	648	2 A25047	beta-glucuronidase
5	879	29.3	603	1 GBEGCG	beta-glucuronidase
6	865.5	28.8	651	2 A26581	beta-glucuronidase
7	691	23.0	368	2 A85768	partial beta-D-glu
8	691	23.0	370	2 D90919	beta-D-glucuroni
9	487	16.2	755	2 D95842	probable beta-gal
10	482	16.1	716	2 JU0275	beta-galactosidase
11	481.5	16.0	996	2 D86872	beta-galactosidase
12	452.5	15.1	1087	2 F72283	beta-galactosidase
13	414	13.8	1024	1 GBEC	beta-galactosidase
14	405.5	13.5	1024	2 E90678	beta-D-galactosida
15	405.5	13.5	1024	2 A85529	beta-D-galactosida
16	381.5	12.7	1014	2 C83990	beta-galactosidase
17	371.5	12.4	1307	2 T35944	probable beta-gala
18	364	12.1	897	2 A39405	beta-galactosidase
19	360.5	12.0	1060	2 A10201	beta-galactosidase
20	352.5	11.7	1025	1 JC1266	beta-galactosidase
21	352.5	11.7	1042	1 GBEGC	beta-galactosidase
22	352.5	11.7	1042	2 E85968	evolved beta-D-gal
23	352.5	11.7	1042	2 F91123	evolved beta-D-gal
24	348.5	11.6	1034	2 T30574	beta-galactosidase
25	348	11.6	1026	2 A49750	beta-galactosidase
26	339.5	11.3	1034	2 A24925	beta-galactosidase
27	328	10.9	1075	2 T47603	beta Galactosidase
28	327.5	10.9	1034	2 T30551	beta-galactosidase
29	322.5	10.7	626	2 A42891	beta-galactosidase

30	321	10.7	2228	2 E97942	beta-galactosidase
31	321	10.7	2233	2 B95075	beta-galactosidase
32	312	10.4	1015	2 I39697	beta-galactosidase
33	310	10.3	1005	2 T31333	beta-galactosidase
34	300.5	10.0	1007	2 A30093	beta-galactosidase
35	256	8.5	785	2 H72228	hypothetical prote
36	194	6.5	237	2 B85768	interrupted beta-D
37	194	6.5	237	2 B85768	partial beta-D-glu
38	191.5	6.4	827	2 AB2764	beta-mannosidase p
39	191.5	6.4	832	2 A97545	hypothetical prote
40	181	6.0	820	2 T37230	probable beta-mann
41	179.5	6.0	900	2 T19689	hypothetical prote
42	179	6.0	879	2 A55881	beta-mannosidase (
43	175	5.8	891	2 A82755	beta-mannosidase p
44	139.5	4.6	1091	2 F83928	hypothetical prote
45	132	4.4	818	2 A97668	mannosidase AGR_C

ALIGNMENTS

RESULT 1									
A72300									
beta-glucuronidase - Thermotoga maritima (strain MSB8)									
C/Species: Thermotoga maritima									
C/Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 09-Jul-2004									
C/Accession: A72300									
R/Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.; C.M.									
Nature 399, 323-329, 1999									
A/Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq									
A/Reference number: A72200; MUID:99287316; PMID:10360571									
A/Accession: A72300									
A/Status: preliminary									
A/Molecule type: DNA									
A/Residues: 1-563 <ARN>									
A/Cross-references: UNIPROT:Q9X0F2; GB:AE001766; GB:AE000512; NID:G4981600; PIDN:AAD3614									
A/Experimental source: strain MSB8									
C/Genetics:									
A/Gene: TM1062									
C/Superfamily: beta-glucuronidase									
Query Match									
Best local Similarity 99.8%; Pred. No. 4.2e-197;									
Matches 562; Conservative 0; Mismatches 1; Indels 0; Gaps 0;									
QY	1	MVRPQRNKKRFFILLINGVWNLEVTSKDRPIAVGSMNEQYQDICYEEGPFYKTFYVPK	60						
DB	1	MVRPQRNKKRFFILLINGVWNLEVTSKDRPIAVGSMNEQYQDICYEEGPFYKTFYVPK	60						
QY	61	XLQKHIRLYFAVNTDCEVFLNGEKGVENHIEYLPEFVDVTGKVKSGENELRVVENRL	120						
DB	61	ELSQKHIRLYFAVNTDCEVFLNGEKGVENHIEYLPEFVDVTGKVKSGENELRVVENRL	120						
QY	121	KVGFPPSKVPDSGTHTVGFFGSPPPANFDFPPYGGIIRPVLIJEFTDHARILDIWDTSES	180						
DB	121	KVGFPPSKVPDSGTHTVGFFGSPPPANFDFPPYGGIIRPVLIJEFTDHARILDIWDTSES	180						
QY	181	EBPEKLGKVKVKEIVSEAVGQEMTIKGEBEKKIRTSNRFEGEFILLENARFWSLEDPY	240						
DB	181	EBPEKLGKVKVKEIVSEAVGQEMTIKGEBEKKIRTSNRFEGEFILLENARFWSLEDPY	240						
QY	241	LYPLKVELKEDEYTLDIGIRTSWDEKRLYLNGKPVFLKFGKHKEEPVLGOGTFYPLMI	300						
DB	241	LYPLKVELKEDEYTLDIGIRTSWDEKRLYLNGKPVFLKFGKHKEEPVLGOGTFYPLMI	300						
QY	301	KDFNLKWINANSFRTSHYPYSEEWLDLADRLGILVIDEAPHYGITRHYHNPETQKIAED	360						
DB	301	KDFNLKWINANSFRTSHYPYSEEWLDLADRLGILVIDEAPHYGITRHYHNPETQKIAED	360						
QY	361	NIRRMIDRHKNHPSVIMSVANEPESNHPDAGFGFKALYETANEMDRTRPVVWVSMADAP	420						

```

Db      361 NRRMIDRHKHNPVIMSVANEPESNHPDAEGFEKALYETANEMDRTRPVVWVSMDAP 420
QY      421 DERTRDVALKYFDIVCNRYGWIYQRIEEGLQALEKDIIEELVARHRKPIFVTEFGAD 480
Db      421 DERTRDVALKYFDIVCNRYGWIYQRIEEGLQALEKDIIEELVARHRKPIFVTEFGAD 480
QY      481 AIAIGHYDPQPMSESEYQALVEKTRILLKKDYIIGTHWAFADFKTPQVRRPILNHK 540
Db      481 AIAIGHYDPQPMSESEYQALVEKTRILLKKDYIIGTHWAFADFKTPQVRRPILNHK 540
QY      541 GVETRDROPKLVAVHLRLMSEV 563
Db      541 GVETRDROPKLVAVHLRLMSEV 563

```

## RESULT 2

```

beta-glucuronidase (gusb) [imported] - Sulfolobus solfataricus
C/Species: Sulfolobus solfataricus
C/Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 09-Jul-2004
C/Accession: C90485
R/She, Q./Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chan-
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, F.
arrett, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.
submitted to GenBank, April 2001
A/Description: Sulfolobus solfataricus complete genome.
A/Reference number: A99139
A/Accession: C90485
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-570 <KUR>
A/Cross-references: UNIPROT:Q97U11; GB:AE006641; NID:g13816434; PIDN:AAK43138.1; GSPDB:C
C/Genetics:
A/Gene: gusb
C/Superfamily: beta-glucuronidase

```

```

Query Match      33.7%; Score 1011; DB 2; Length 570;
Best Local Similarity 39.2%; Pred. No. 2, 7e-61;
Matches 230; Conservative 92; Mismatches 191; Indels 74; Gaps 14;

QY      15 LNVGNVLEVTSKDRP-----IAVPGSWNEQYQDLCYEEGPPTKTTFYVPK 60
Db      11 LQGFWMFKINDENTGEENGWYKGLESEDIYVPASWNEQNPCKWDQFSGIAYQKDLFVSN 70

QY      61 XLSQKHRLYFAAVNTDCEVFLNGEKVGENHIEYLPFEVDVTGKVKSGENELRVVVENRL 120
Db      71 DNGNRKAMWVFEAGAYITKLMINGEGYGTGHSFTQFKPKIKLV---NEFNKIV---V 123

QY      121 KVGGFPSKVDSTGTHVGFSGSPPPAN-----FDFFPYGGIIRPVLIIEFTDHARILDIW 174
Db      124 KIDNTPSPY-----NLPPARDLNNAAFDFPNYGGIHRPVYIEFVDECHVEDIT 171

QY      175 VDTSESEPEKKLGKVKKIEVSEAVGQEMTIKLGEEKKI---RTSNRFVEGEPIIEN 230
Db      172 VYT-----KSYGHLKVEI-LSECNQRFSLRFLVDKGRVITLNESSNEVFEKD--VNN 222

QY      231 ARFWSLEDYLYPLKVELE---KDEYTLDIGIRTSWDEKRLYLNKRPVFLKGFGRHE 285
Db      223 VIPWSPDNLYLTLIVEMVGGNLKDSVYERIGFRDVEYKDKIYLNGKPIFLKGFGRHE 282

QY      286 EEPVLGGCTFFPLMIKDFNLKMINANSRTSHYPYSEEWLADRLGILVIDEAP--HV 343
Db      283 DFPILGKFTYGAVALVRDFPLMRKIGANSFRTSHYPYSEHLDLADENGFLVILEPLCY 342

QY      344 GTRHYHNETOKI-----AEDNTRMIDRHKHNPVIMSVANEPESNHPDAEGF 394
Db      343 NISRVMSGEELAKMFGDVYKFEKVRDTIKEMIRQHKRPSVIMYSVWNEPPSDIREVAEF 402

QY      395 FKALYETANEMDRTRPVVWVSMDAPDERTRDVALKYFDIVCNRYGWIYQRIEEGL 454
Db      403 IRREVELEKSLDSSRPVTFAS-----HRSVRDLALEYDVISLNYHGMWTEWGDIDSGV 457

QY      455 QALEKDIIEELVARH-RKIFVTEFGADAIAGIHYDPQPMSESEYQALVEKTRILLKKD 513

```

```

Db      458 KVAIELEIEIHKKPFPEKPIITTEFGADAIYGLHSDPQMWSESEYQSEMIRKYIEALREKD 517
QY      514 YIIGTHWAFADFKTPQVRRPILNHKGVETRDROPKLVAVHLRLW 560
Db      518 YIVGFHWNFADFRTPQNPSTRITLNKGIIFTRDRQPKLAQVVEELF 564

```

## RESULT 3

```

A32576
beta-glucuronidase (BC 3.2.1.31) allele B precursor - mouse
N/Alternate names: beta-D-glucuronoside glucuronosohydrolase
C/Species: Mus musculus (house mouse)
C/Date: 12-Oct-1989 #sequence_revision 12-Oct-1989 #text_change 09-Jul-2004
C/Accession: A32576; B32576; I49692; A28954; A29977; A35798
R/Wawrzyniak, C.J.; Gallagher, P.M.; D'Amore, M.A.; Carter, J.E.; Lund, S.D.; Rinchik, E.
Mol. Cell. Biol. 9, 4074-4078, 1989
A/Title: DNA determinants of structural and regulatory variation within the murine beta-g
A/Reference number: A32576; MUID:89384641; PMID:2779578
A/Accession: A32576
A/Molecule type: mRNA
A/Residues: 1-648 <WAW>
A/Cross-references: UNIPROT:P12265; GB:M28540; GB:M27816; NID:g193718; PIDN:AAA63307.1; I
A/Experimental source: allele B
A/Accession: B32576
A/Molecule type: mRNA
A/Residues: 1-86, 'I', 88-648 <WA2>
A/Cross-references: GB:M28541; NID:g193720; PIDN:AAA63308.1; PID:g193721; GB:M27816
A/Experimental source: allele H
R/Funkenstein, B.; Leary, S.L.; Stein, J.C.; Catterall, J.F.
Mol. Cell. Biol. 8, 1160-1168, 1988
A/Title: Genomic organization and sequence of the Gus-s-a allele of the murine beta-gluci
A/Reference number: I49692; MUID:88216590; PMID:2835664
A/Accession: I49692
A/Status: translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-232, 'T', 234-264, 'D', 266-319, 'V', 321-427, 'K', 429-615, 'L', 617-648 <RES>
A/Cross-references: GB:M19279; NID:g193524; PIDN:AAA37697.1; PID:g309257
A/Experimental source: allele A
R/D'Amore, M.A.; Gallagher, P.M.; Korfhagen, T.R.; Ganschow, R.E.
Biochemistry 27, 7131-7140, 1988
A/Title: Complete sequence and organization of the murine beta-glucuronidase gene.
A/Reference number: A28954; MUID:89062453; PMID:3196706
A/Accession: A28954
A/Molecule type: DNA
A/Residues: 1-264, 'D', 266-319, 'V', 321-648 <DAM>
A/Cross-references: GB:J02836; GB:J03035; GB:M20204; NID:g193716; PIDN:AAA98623.1; PID:g:
R/Gallagher, P.M.; D'Amore, M.A.; Lund, S.D.; Ganschow, R.E.
Genomics 2, 215-219, 1988
A/Title: The complete nucleotide sequence of murine beta-glucuronidase mRNA and its deduc
A/Reference number: A29977; MUID:88284700; PMID:3397060
A/Accession: A29977
A/Molecule type: mRNA
A/Residues: 1-264, 'D', 266-319, 'V', 321-648 <GAL>
A/Cross-references: GB:J03047; NID:g193522; PIDN:AAA37696.1; PID:g309256
R/Li, H.; Takeuchi, K.H.; Manly, K.; Chapman, V.; Swank, R.T.
J. Biol. Chem. 265, 14732-14735, 1990
A/Title: The propeptide of beta-glucuronidase. Further evidence of its involvement in cor
A/Reference number: A35798; MUID:90368633; PMID:2394691
A/Accession: A35798
A/Status: not compared with conceptual translation
A/Molecule type: mRNA
A/Residues: 593-648 <LIA>
A/Note: the location of the propeptide cleavage site was not demonstrated directly but re
C/Comment: In some tissues, a portion of this enzyme is retained in the endoplasmic retic
syn.
C/Genetics:
A/Gene: Gus
A/Map position: 5
A/Intons: 70/3; 132/3; 193/2; 241/1; 303/3; 351/3; 411/2; 460/2; 488/3; 547/3; 593/1
C/Superfamily: beta-glucuronidase
C/Keywords: glycosidase; hydrolase; lysosome

```







```
Db      147 LPFEADISNLVQVGPLPSRLRITIAINNTLT-----PTTLPGTIGYLTDTSKYKGYF-- 200
QY      143 PPANPDFFPYGGIIRPVLIETDARILDIWDTSESEPEKKLGKVKKIEVEEAVGQ 202
Db      201 VQNTYFDFFNAYAGIQRSVLLYTTPTTYIDITVTTSS---VEQDSGLVNYQISVKGSNL-F 256
QY      203 EMTIKLGEEKKIRTSNRFVEGEFIIENARFW----SLEDP-VLYPLKVELE----- 249
Db      257 KLEVRLLDAENKVVANGTGTGQGLKPGVSLMWPMYLMHERPALYLSLEVQLTAQTSLGPV 316
QY      250 KDEYTLDIGIRTSWDEKRLYLNGKPVFLKGFGEFPVLGGGTFFPLMIKDFNLKMI 309
Db      317 SDFYTLPGIRTAVTKSGQLINGKRFYFHGVNKGEDADIRKGEDWPLVKDFNLLRWL 376
QY      310 NANSFRTSHYPYSEEWLADRLGLVIDEAPHVG-----ITRYHNPETQKIAED 360
Db      377 GANAFTSHYPYAEVWMQMDRGIVVIDECPGVGLALPQFPNNVSLHH-----MQVME 432
QY      361 NIRMIDRHKNHPSVIMMSVANEPESNHPDAEGFFKALYETANEMDRTPVVVMSMDAP 420
Db      433 VVR-----RDKNHAVVMMSVANEPASHLESAGYLLKAVIAHTKSLDPSRPVTFVS---N 484
QY      421 DERTRDVALKYPDIVCVNRYYGWYIQGRIEGLQALEKDIELLYARHKRPITPEFGAD 480
Db      485 SNYAADKGAPYDVICLNSYSYSHYDYGHLLELIQLATQFENMYKKYOKPIITSEYCAE 544
QY      481 AIAGIHYDPQPMSESEYQAEIVETIRLL--KKDYIIGTHVAFADFKTPQNVRRPIL 537
Db      545 TIAGFHQDPPLMFTEEYQKSLER- YHGLDQKRKYVVGELIWNFADFMTESPTRVLG 603
QY      538 NHKGVFTRDRQPKLVAVHLR-RLW 560
Db      604 NKKGIFTRQRPKSAAPFLIRERYW 627
```

```
RESULT 7
A85768 partial beta-D-glucuronidase [imported] - Escherichia coli (strain O157:H7, substrain ED
C/Species: Escherichia coli
C/Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 09-Jul-2004
C/Accession: A85768
R/Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
Hler, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A/Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A/Reference number: A85480; MUID:21074935; PMID:11206551
A/Accession: A85768
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-368 <STO>
A/Cross-references: UNIPROT:Q8X671; GB:AE005174; NID:g12515602; PIDN:AAG56605.1; GSPDB:C
A/Experimental source: strain O157:H7, substrain EDL933
C/Genetics:
A/Gene: uidA_1
```

```
Query Match      23.0%; Score 691; DB 2; Length 368;
Best Local Similarity 39.6%; Pred. No. 1e-39;
Matches 141; Conservative 62; Mismatches 125; Indels 28; Gaps 4;

QY      230 NARFWSLEDPLYPL---KVELEKDEYTLDIGIRTSWDEKRLYLNGKPVFLKGFGEKHE 285
Db      3 NPHLMQPGEGYLYELVYTAKSRTCEDIYPLRVGIRSAVAVKGEQFLINHKKPFYFGFGRHE 62
QY      286 EFPVLGGGTFFPLMIKDFNLKMINANSFRTSHYPYSEEWLADRLGILVIDEAPHVGI 345
Db      63 DADLRGKGFQNVLMVHDHALMDWIGANSYRTSHYPAEEMLDMADEHGIVIDEATAVGF 122
QY      346 -----TRYHNPETQKIAEDNIRRMIDRHKNHPSVIMMSVANEPESN 387
Db      123 NLSLIGIFEAGNKPELYSEEAVENTQQAHLQAIKELIARDKNHPSVMMSTIANEPDTR 182
QY      388 HPDABGFPAKALYETANEMDRTPVVVMSM--DAPDERTDVALKYFDIVCVNRYYGWYI 445
```

```
Db      183 PGABEYFAPPLAEATRKLDPTRITTCVNVMFCDAHTDTISDL-----FDVLCLNRYYGWYV 238
QY      446 YQGRIEEGLQALEKDIELVYARHKRPVFVTEFGADAIAGIHYDPQPMFSESEYQAEVEXT 505
Db      239 QSGDLETAEKYLEKEILLAWQEKHQPIITTEYGVDTLAGLHSMYTDMMSESEYQCAWLDMY 298
QY      506 IRLLLKDYITIGTHVAFADFKTPQNVRRPILNHKGVFTRDRQPKLVAVHLRLWS 561
Db      299 HRVFDVSAVVEQVWNPADPATSGILRVGNNKKGIFTRDRKPKSAAPFLQKRWY 354
```

```
RESULT 8
beta-D-glucuronidase [imported] - Escherichia coli (strain O157:H7, substrain RIMD 05099
D90919
C/Species: Escherichia coli
C/Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 09-Jul-2004
C/Accession: D90919
R/Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A/Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno
A/Reference number: A99629; MUID:21156231; PMID:11258796
A/Accession: D90919
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-370 <HAY>
A/Cross-references: UNIPROT:Q8X671; GB:BA000007; PIDN:BA835747.1; PID:g13361791; GSPDB:C
A/Experimental source: strain O157:H7, substrain RIMD 0509952
C/Genetics:
A/Gene: ECs2324
```

```
Query Match      23.0%; Score 691; DB 2; Length 370;
Best Local Similarity 39.6%; Pred. No. 1e-39;
Matches 141; Conservative 62; Mismatches 125; Indels 28; Gaps 4;

QY      230 NARFWSLEDPLYPL---KVELEKDEYTLDIGIRTSWDEKRLYLNGKPVFLKGFGEKHE 285
Db      5 NPHLMQPGEGYLYELVYTAKSRTCEDIYPLRVGIRSAVAVKGEQFLINHKKPFYFGFGRHE 64
QY      286 EFPVLGGGTFFPLMIKDFNLKMINANSFRTSHYPYSEEWLADRLGILVIDEAPHVGI 345
Db      65 DADLRGKGFQNVLMVHDHALMDWIGANSYRTSHYPAEEMLDMADEHGIVIDEATAVGF 124
QY      346 -----TRYHNPETQKIAEDNIRRMIDRHKNHPSVIMMSVANEPESN 387
Db      125 NLSLIGIFEAGNKPELYSEEAVENTQQAHLQAIKELIARDKNHPSVMMSTIANEPDTR 184
QY      388 HPDABGFPAKALYETANEMDRTPVVVMSM--DAPDERTDVALKYFDIVCVNRYYGWYI 445
Db      185 PGABEYFAPPLAEATRKLDPTRPITCVNVMFCDAHTDTISDL-----FDVLCLNRYYGWYV 240
QY      446 YQGRIEEGLQALEKDIELVYARHKRPVFVTEFGADAIAGIHYDPQPMFSESEYQAEVEXT 505
Db      241 QSGDLETAEKYLEKEILLAWQEKHQPIITTEYGVDTLAGLHSMYTDMMSESEYQCAWLDMY 300
QY      506 IRLLLKDYITIGTHVAFADFKTPQNVRRPILNHKGVFTRDRQPKLVAVHLRLWS 561
Db      301 HRVFDVSAVVEQVWNPADPATSGILRVGNNKKGIFTRDRKPKSAAPFLQKRWY 356
```

```
RESULT 9
probable beta-galactosidase (EC 3.2.1.23) [imported] - Sinorhizobium meliloti (strain 10
D95842
C/Species: Sinorhizobium meliloti
C/Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 09-Jul-2004
C/Accession: D95842
R/Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Hernan
Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001
A/Title: The complete sequence of the 1,683-kb pSymb megaplasmid from the N2-fixing endo
A/Reference number: A95842; MUID:21396508; PMID:11481431
A/Accession: D95842
A/Status: preliminary
```

A/Molecule type: DNA  
A/Residues: 1-755 <KUR>  
A/Cross-references: UNIPROT:Q92XF7; GB:AL591985; PIDN:CAC48404.1; PID:g15139876; GSPDB:C  
A/Experimental source: strain 1021, megaplasmid pSymB  
R/Gallbert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,  
pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;  
L.; Hyman, R.W.; Jones, T.  
Science 293, 668-672, 2001  
A/Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Leleure,  
hebaull, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.  
A/Title: The composite genome of the legume symbiont *Sinorhizobium meliloti*.  
A/Reference number: A96039; MUID:21368234; PMID:11474104  
A/Contents: annotation  
C/Genetics:  
A/Gene: lacZ1; SMD21655  
A/Genome: plasmid  
C/Superfamily: beta-galactosidase  
C/Keywords: glycosidase; hydrolase

C/Accession: D86872  
R/Bolotin, A.; Winkler, P.; Mauger, S.; Jallou, O.; Malarme, K.; Weissenbach, J.; Enlth  
Genome Res. 11, 731-753, 2001  
A/Title: The complete genome sequence of the lactic acid bacterium *Lactococcus lactis* s  
A/Reference number: A86625; MUID:21235186; PMID:11337471  
A/Accession: D86872  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-996 <STO>  
A/Cross-references: GB:AE005176; PID:G12725024; PIDN:AAK06078.1; GSPDB:GN00146  
A/Experimental source: strain IL1403  
C/Genetics:  
A/Gene: lacZ  
C/Superfamily: beta-galactosidase  
C/Keywords: glycosidase; hydrolase

Query Match	16.0%;	Score 481.5;	DB 2;	Length 996;
Best Local Similarity	26.3%;	Pred. No. 8.7e-25;		
Matches 162; Conservative	91;	Mismatches 209;	Indels 153;	Gaps 23;

```

OY 15 LNWGVN-----LEVTSKDRPIAVPGSWNEOYODLCYEKGPFYKTF---- 56
    ||: ||      ||: ||      ||: ||      ||: ||
Db 38 LNLGMNFDHFSRISDVPKNWLELTESKTEIIVPSNMQIEFKD--KSDVPIYTNVTYPIPI 95
OY 57 ---YVPKX-----LSQKHRLYPAAVNTDCEVFLNGEKVGENHIEYLPF 97
    ||: ||      ||: ||      ||: ||      ||: ||
Db 96 QPYPVPEANPVGAYSRYFDITKEWLESGHVHLTFEVGSAFHFWLNGEYGSSEDSRLPA 155
OY 98 EVDVTGKVKSGENELRVVVENRLKVGGEPPSKVPDSCHTVGFPGSFPPANPDFPPIYGGII 157
    ||: ||      ||: ||      ||: ||      ||: ||
Db 156 EFDISNLAKEGONCLKVLVFRWSKVITYFEDQ-----DMWRMSGIF 195
OY 158 RPLVIEFTDHARIDIVIDTSESEPEKKLGKVKKIEVSEBAVGQE-WIKLGEEEKR 216
    ||: ||      ||: ||      ||: ||      ||: ||
Db 196 RSNVLQWLPDNYLLDFSIXT--DLDEDLPANVKQAYAKNIDDACLEFKLYDDEQLIG 252
OY 217 TSNRFVEGEFILENARFWSLEDPLYPLKVELEK-----DEYTLDIGRTISWDEKRLY 270
    ||: ||      ||: ||      ||: ||      ||: ||
Db 253 ECHGF-DAEIGVNPXKLMSDEIPYLRYLELTLMDRSGAVFHKETKIGIRKIAIEKGQLK 311
OY 271 LNKRPVFLKGFGKHEEFVVLGGSTFYPLMIKDFNLKWINANSFRTSHYPYSEEWLIDLAD 330
    ||: ||      ||: ||      ||: ||      ||: ||
Db 312 INGCALLVRGVNGHEFTPEHGVVSEEPWIKIDIKMKENFNNAVRCSHYPNDSRWYELCD 371
OY 331 RLGLLVIDEA---PHYGIT---RHYNPETQKIAEDNTRMIDRHKHNSVIMWSVANEP 384
    ||: ||      ||: ||      ||: ||      ||: ||
Db 372 EYGLIYVMEANIETH-GMTPMNRILTNDPTYLLPLMSERATRMVMRDRNHSIIISWLSGNS 430
OY 385 --ESNHPDAEGFFKALYETANEMDRPVMVMSMDAPDERTRDVALKYFDIVC----- 436
    ||: ||      ||: ||      ||: ||      ||: ||
Db 431 GYGSNH-----QALYDWCKSFDSRPV---HYEGGDASRG-ATDRTDIICPMYARV 478
OY 437 ---VNRYYGWYIYQGRIBEGLQALEKDIEELYARHRKPIFVTEFGAD---AIAIGHYDP 489
    ||: ||      ||: ||      ||: ||      ||: ||
Db 479 DSPSINAPYSLKTWMC-----VAGENRPLILCEYAHDMGNSLSG----- 517
OY 490 PQMFSEBYQAEIVKTRIRLLKKDYYIIGTHVWAF-----ADFKTPQNVRRP 535
    ||: ||      ||: ||      ||: ||      ||: ||
Db 518 ---FGKYWQA-----FREIDRLOGGFIMDWVDOGLLKDGNAYAGDGFQDKPNDRQF 565
OY 536 ILNHKGVFTRDRQPK 550
    ||: ||      ||: ||      ||: ||      ||: ||
Db 566 SLN--GLVFPNRQAK 578

```

RESULT 12  
F72283  
beta-galactosidase - *Thermotoga maritima* (strain MSB8)  
C/Species: *Thermotoga maritima*  
C/Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000  
C/Accession: F72283  
R/Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey  
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.

C.M.  
Nature 399, 323-329, 1999  
A/Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq  
A/Reference number: A72200; MUID:99287316; PMID:10360571  
A/Accession: F72283  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-1087 <ARN>  
A/Cross-references: GB:AE001776; GB:AE000512; NID:g4981740; PIDN:AAD36268.1; PID:g498174  
A/Experimental source: strain MSB8  
C/Genetics:  
A/Gene: TM1193  
C/Superfamily: beta-galactosidase

Query Match	15.1%;	Score 452.5;	DB 2;	Length 1087;
Best Local Similarity	26.2%;	Pred. No. 9.4e-23;		
Matches 163; Conservative	97;	Mismatches 222;	Indels 141;	Gaps 25

QY	15	LVGNWN-----LEV-----TSKDRPIAVGSGWNEQ-----YQDLCY---EEG	48
Db	45	LVGNWRFLFAKNPEVPEPDEFSEKFPDSNWDEIEVBSNWEMKGYGKPIYTNVVPPEPNP	104
QY	49	PF-----TYKTFYVPKLSQKHIRLYFAAVNTDCEVFLNGEKYGENHIEYLPFEV	99
Db	105	PFVPKDDNPTGVYRRWIEIPEDWEKKEIFLHFEGVRSFFYLWVNGKKGIFS KDSCTPAEF	164
QY	100	DVTGKVKSGENELRVVVENRLKVGGFPSKVPDSCGHTVGFPGSFPPANPDFFPYGGIIRP	159
Db	165	RLTDVLRPGKNLITVEV-----LKMSD-----GSY-LEQDMMWFAGIYRD	204
QY	160	VLEIFTDHARLIDIVDTSSESEPEKLGKVKVKEIVSEBAVGQEMTIKLGEEBK----	214
Db	205	VLYLALPKPHIRDVFRVTRDLDEYNRN-GKIFLDVENR-----NLGEEEEKDLEVT	253
QY	215	-----IRTSNRFVEGEPFILENARFWSLEDPLYPLKYVLEKDEYTLDIGI	259
Db	254	LITPDGDEKTLVKEITVKPEDRVLSFAFDVKDPKKWSAETPLHYLVKLKLGEDKVKVNGF	313
QY	260	RTISWDEKRLYLVNGKPFVFLKGFSGHSEFPVLGGGTFYPLMIKDFNLKMINANSFRTSHY	319
Db	314	RKIEIKDGTLLFMGKPLYIKGVNRHBFDPDRGHAVTVERMIQDIKLMKQHNINTVRTSHY	373
QY	320	PYSEEWLDLADRLGILVIDEAPHVIGITRYHNPEF-----QKIAEDNIRRMIDRHKN	371
Db	374	PNQTKMYDLCDFGLYVIDEA-NIESHGIDWDPEVTLANRWEMEKAHDRIKRMVERDKN	432
QY	372	HPSVIMWSVANEPESNHPDAEGFPKALYETANEMDTRPVVMVSMMDAPDERTRDVALKY	431
Db	433	HPSIIFWSLIGNEA---GDGVNFEXAAL-WIKKRONTRLIHY-----EGTTRRGESY	480
QY	432	FDIVCVNRRYYGYIYQGRIEEGLQALEKDIIEELVA--RHRKPIFVTEFG--ADAIAGI-	485
Db	481	VDVF-----SLMYPKMDILLEVASKKREKPFIMCEYAHAMGNSVGNLK	523
QY	486	-HYDRPQMFSEEYQA---ELVEKTRILLKDYIIGTHWAF-ADFKTPOQNRVRPILNKH	540
Db	524	DYWDVIEKYPYLLHGGCICIMDWVDQIR--KKDE-NGREFWAYGDFGDTPNDFNFCIN--	577
QY	541	GVFTRDRQPKLVAHVLRLWSEV	563
Db	578	GVVLPDRTPPEPELYEVKKVYQNV	600

RESULT 13  
GBEC  
beta-galactosidase (EC 3.2.1.23) lacZ [validated] - Escherichia coli (strain K-12)  
N;Alternate names: beta-D-galactosidase; lactase  
C;Species: Escherichia coli  
C;Date: 24-Apr-1984 #sequence\_revision 23-Sep-1997 #text\_change 09-Jul-2004  
C;Accession: A90981; A92233; A93224; S06878; I41218; H64761; I40987; A00898; S14637; S14  
R;Kalinins, A.; Otto, K.; Ruthner, U.; Muller-Hill, B.  
EMBO J. 2, 593-597, 1983  
A;Title: Sequence of the lacZ gene of Escherichia coli.

A/Reference number: A90981; MUID:84028567; PMID:6313347  
 A/Accession: A90981  
 A/Molecule type: DNA  
 A/Residues: 2-1024 <KAL>  
 A/Cross-references: UNIPROT:P00722; GB:V00296; NID:g41901; PIDN:CAA23573.1; PID:g1197203  
 A/Note: translation of initiator Met is not shown  
 R/Fowler, A.V.; Zabin, I.  
 J. Biol. Chem. 253, 5521-5525, 1978  
 A/Title: Amino acid sequence of beta-galactosidase. XI. Peptide ordering procedures and  
 A/Reference number: A92233; MUID:78218239; PMID:97298  
 A/Accession: A92233  
 A/Molecule type: protein  
 A/Residues: 2-1024 <POW>  
 A/Note: this is the final paper in a series  
 R/Calos, M.P.; Miller, J.H.  
 Nature 285, 38-41, 1980  
 A/Title: Molecular consequences of deletion formation mediated by the transposon Tn9.  
 A/Reference number: A93224; MUID:80188189; PMID:6246435  
 A/Accession: A93224  
 A/Molecule type: DNA  
 A/Residues: 356-476 <CAL>  
 R/Rutshouser, E.C.; Richardson, J.P.  
 J. Mol. Biol. 208, 23-43, 1989  
 A/Title: Identification and characterization of transcription termination sites in the E  
 A/Reference number: S06878; MUID:89362462; PMID:2475637  
 A/Accession: S06878  
 A/Status: translation not shown  
 A/Molecule type: mRNA  
 A/Residues: 1-147 <RUT>  
 A/Cross-references: EMBL:X16313; NID:g41903; PIDN:CAA34380.1; PID:g41904  
 R/Mikryukov, N.N.; Petrov, N.A.; Karginov, V.A.; Vassilenko, S.K.  
 Biorg. Khim. 6, 1735-1736, 1980  
 A/Title: Nucleotide sequence of a lambda-plac 5-1 DNA region coding for a COOH-terminal  
 A/Reference number: I41218  
 A/Accession: I41218  
 A/Status: translated from GB/EMBL/DBJ  
 A/Molecule type: DNA  
 A/Residues: 967-971, 'R', 973-1022, 'E', 1024 <MIK>  
 A/Cross-references: GB:X38327; NID:g146061; PIDN:AAA23835.1; PID:g146062  
 R/Jacobson, R.H.; Zhang, X.; Dubose, R.F.; Matthews, B.W.  
 submitted to the Brookhaven Protein Data Bank, July 1994  
 A/Reference number: A65162; PDB:1BGL  
 A/Contents: annotation; X-ray crystallography, 2.50 angstroms, residues 4-1024  
 R/Jacobson, R.H.; Zhang, X.J.; Dubose, R.F.; Matthews, B.W.  
 Nature 369, 761-766, 1994  
 A/Title: Three-dimensional structure of beta-galactosidase from E. coli.  
 A/Reference number: A58594; MUID:94277211; PMID:8008071  
 A/Contents: annotation; X-ray crystallography, 2.50 angstroms  
 R/Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co  
 .A.; Rose, D.J.; Mau, B.; Shao, Y.  
 Science 277, 1453-1462, 1997  
 A/Title: The complete genome sequence of Escherichia coli K-12.  
 A/Reference number: A64720; MUID:97426617; PMID:9278503  
 A/Accession: H64761  
 A/Status: nucleic acid sequence not shown; translation not shown  
 A/Molecule type: DNA  
 A/Residues: 1-1024 <BLAT>  
 A/Cross-references: GB:AE000141; GB:U00096; NID:g1786532; PIDN:AACT3447.1; PID:g1786539;  
 A/Experimental source: strain K-12, substrain MG1655  
 R/Prencki, P.  
 Gene 122, 231-232, 1992  
 A/Title: Nucleotide sequence of the classical lacZ deletion delta M5.  
 A/Reference number: I40987; MUID:93083990; PMID:1339377  
 A/Accession: I40987  
 A/Molecule type: DNA  
 A/Residues: 1-11, 43-50 <RES>  
 A/Cross-references: EMBL:X58252; NID:g40882; PIDN:CAA41206.1; PID:g40883  
 C/Genetics:  
 A/Gene: lacZ  
 A/Map position: 8 min  
 C/Complex: homotetramer  
 C/Function:  
 A/Description: catalyzes hydrolysis of lactose into galactose and glucose

C/Superfamily: beta-galactosidase  
 C/Keywords: glycosidase; homotetramer; hydrolase; magnesium  
 F/2-1024/Product: beta-galactosidase #status experimental <MAT>  
 F/2-50/Region: alpha complementation  
 F/51-218/Domain: 1, jelly-roll beta-barrel #status predicted <DM1>  
 F/219-334/Domain: 2, fibronectin type-III fold #status predicted <DM2>  
 F/335-627/Domain: 3, distorted TIM barrel #status predicted <DM3>  
 F/628-737/Domain: 4, fibronectin type-III fold #status predicted <DM4>  
 F/738-1024/Domain: 5, anti-parallel beta-sandwich #status predicted <DM5>  
 F/417, 419, 462/Binding site: magnesium (Glu, His, Glu) #status experimental  
 F/462, 504, 538/Active site: Glu, Tyr, Glu #status predicted

Query Match 13.8%; Score 414; DB 1; Length 1024;  
 Best Local Similarity 28.4%; Pred. No. 3.7e-20;  
 Matches 132; Conservative 59; Mismatches 172; Indels 102; Gaps 16;

Qy	15	LNQVNVLEVTSKDRPIAVPGSWNE-----QYQDLQYB---EG	48
Db	55	LNQEWRF--AMFPAPAVPESWLECDLPEADTVVPSNMQMHGYDAPITYNTVPIYVNP	112
Qy	49	PFT-----YKTFEVPKX-LSQKHIRLYFAAVNTDCEVFLNGEKVGENHIEYLPFEV	99
Db	113	PFTPTNPPTGCSLTFENVDESWLQEQTRIIFDGVNSAFHLMCNGWVGXQDSRLPSEF	172
Qy	100	DVTGKVGSENELRVVVENRLKVGEPSPKVPDSGTHVGFSGFPANPDFPFGGIRP	159
Db	173	DLAPLRAGENRLAVVW-LRWSDGSY---LEDQ-----DMWRMSGIFRD	212
Qy	160	VLIETDARILDIWDTSESEPEKKLGKVKVIEVSEAVGQ-----EMTIKLEEEK	214
Db	213	VSLHKPTQISDFHATRFNDPS----RAVLAEVQMGELRDYLRVTVSLWGSETQ	267
Qy	215	IRTSNRFVEGEFT-----LENARFMSLEDPLYLPVLEKEDEYTL----	255
Db	268	VASGTAPFGGEIIDERGGYADRVTLRLNVENPKLMSAEIPNLVRAVVELHTADGTLIEAE	327
Qy	256	--DIGRTISWDEKRLYLNGKPVPLKGFGEKHEEPVVGQTFYPLMIKDFNLKWINANS	313
Db	328	ACDVGFREVRIENGLLLNGKPLRLGVNRHHEHPLHGQVMDEQTVQDILLMKONFNA	387
Qy	314	FRTSHYPYSEEWLADRLGILVIDEA--PH--VGITRYHNPETQKIAEDNIRRMIDR	368
Db	388	VRCSHPNHPMLWYTLCDRYGLVYVDEANIEHGMVPMNRLTDDPRWLPAWSERVYTMVQR	447
Qy	369	HKMHPVIMWSVANEP--ESNHPDAEGFKALETANEMDRTPV	411
Db	448	DRNHSVITWISLNGESGHANH-----DALYRWIKSVDPSPRV	485

RESULT 14  
 E90678  
 beta-D-galactosidase [imported] - Escherichia coli (strain O157:H7, substrain RIMD 05099;  
 C/Species: Escherichia coli  
 C/Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 09-Jul-2004  
 C/Accession: E90678  
 R/Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;  
 Gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
 DNA Res. 8, 11-22, 2001  
 A/Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genom  
 A/Reference number: A99629; MUID:21156231; PMID:11258796  
 A/Accession: E90678  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-1024 <HAY>  
 A/Cross-references: UNIPROT:Q8X685; GB:BA000007; PIDN:BA83820.1; PID:g13359854; GSPDB:GN  
 A/Experimental source: strain O157:H7, substrain RIMD 0509952  
 C/Genetics:  
 A/Gene: E90678  
 C/Superfamily: beta-galactosidase

Query Match 13.5%; Score 405.5; DB 2; Length 1024;  
 Best Local Similarity 25.2%; Pred. No. 1.4e-19;  
 Matches 159; Conservative 88; Mismatches 216; Indels 169; Gaps 26;



```

QY 15 LNVGNVLEVTSKDRPIAVPGSWNE-----QYODLCYE--EG 48
    |||:| | | | | |
Db 55 LNWEM--QFVWFPAPAEAVPESWLECDLPDADTVVVP SNWOMGTDAPITYNTVYPTVNP 112
    |||:| | | | | |
QY 49 PFT-----YKTFYVPKX-LSOKHIRLYFAAVNTDCEVFLNGEKVGENHIEYLPFEV 99
    |||:| | | | | |
Db 113 PFVPTENPTGCSLTFFNVDESMLQEGOTRIIFDGVNSAFHLMCNGRWVGCGQDSRLSEF 172
    |||:| | | | | |
QY 100 DVTGKVKSGENELRVVVENRLKVGCFPSKVPDSCGTHTVGFFGSPFPANPDFPYGIIIRP 159
    |||:| | | | | |
Db 173 DLSAFLRAGENRLAVMV-LRWSDSY---LEDQ-----DMRMSCIFRD 212
    |||:| | | | | |
QY 160 VLIEFTDHARILDIWVDTSESEPEKKL--GKYVKIEVSEAVQEMTIKLGEEKIR 216
    |||:| | | | | |
Db 213 VSLHKFTTQISDFHVATLFNDBSRVAVLEAVQMYGELRDEL---RTVSLMGETQVA 269
    |||:| | | | | |
QY 217 TSNRFVEGEFI-----LENARFMSLEDPLYLPLKVELKEDEYTL----- 255
    |||:| | | | | |
Db 270 SGTAPEGGEIIDERGGYADRVTLGLNVENPKLMSAEIPNIYRAVELHTADGTLIEAEC 329
    |||:| | | | | |
QY 256 DIGIRTIWDEKRLYLNGKPVFLKGFSGKHEEPVLGQGTFFYPLMIKDNLKWINANSFR 315
    |||:| | | | | |
Db 330 DVGFRVRIENGILLNGKPLLRGVNRHEHHPLHGQVMDQGTQVQDILLMKQNNFNNAV 389
    |||:| | | | | |
QY 316 TSHYPYSEEWLIDLADRLGILVIDEA--PH--VGITRYHNPETOKIAEDNIRMIDRHK 370
    |||:| | | | | |
Db 390 CSHYPNHPLWYTLCDRGLYVVDENANIEHGWPMNRLTDDPWLPMASERATRMVQRDR 449
    |||:| | | | | |
QY 371 NHPSVIMSVANEP--ESNHPDAEGFFKALYETANEMDRTPV-----VMVS 415
    |||:| | | | | |
Db 450 NHPSVIIMSLNESGHGANH-----DALYRWIKSVDPSPRYQEGGADTSATDIIICP 502
    |||:| | | | | |
QY 416 MMDAPDERTRDVALKYFDIVCVNRRYYGYIQGRIEGLQALEKDI BELYARHRKPIFVT 475
    |||:| | | | | |
Db 503 MYARVDEDOFPFPAVPKMSI-----KKWLSLPGEM-----RPLILC 537
    |||:| | | | | |
QY 476 EFG--ADAIAGIHYPDPPQMFSEBYQA-----ELVEKTIRLLKKDYIIGTH 519
    |||:| | | | | |
Db 538 EYAHAMGNSLGG-----FAKYWQAFRQYPRLOGGFVWDLVDQS--LIKYE--NGN 584
    |||:| | | | | |
QY 520 VAAF--ADFKTPQNVRRPILNHKGVFTRDRQP 549
    |||:| | | | | |
Db 585 PMSAYGDFGDTPNDRQFCMN--GLVFADRTIP 614
    |||:| | | | | |

```

RESULT 15  
 A85529  
 beta-D-galactosidase [imported] - *Escherichia coli* (strain O157:H7, substrain EDL933)  
 C/Species: *Escherichia coli*  
 C/Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 09-Jul-2004  
 C/Accession: A85529  
 R/Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew  
 Miller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,  
 Nature 409, 529-533, 2001  
 A/Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.  
 A/Reference number: A85480; MUID:21074935; PMID:11206551  
 A/Accession: A85529  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-1024 <STO>  
 A/Cross-references: UNIPROT:Q8X685; GB:AE005174; NID:G12513175; PIDN:AAG54693.1; GSPDB:Q  
 A/Experimental source: strain O157:H7, substrain EDL933  
 C/Genetics:  
 A/Gene: lacZ  
 C/Superfamily: beta-galactosidase

```

Query Match      13.5%; Score 405.5; DB 2; Length 1024;
Best Local Similarity 25.2%; Pred. No. 1.4e-19;
Matches 159; Conservative 88; Mismatches 216; Indels 169; Gaps 26;

QY    15 LNWGNLEVTSKDRPIAVPGSWNE-----QYDLCYE---EG 48
      ||||| : ||||| : : :
```

Db 55 LNWGEW--QFWWFPAPAEAVPESWLECDLDPADTTVVVPSSNWQMHGYDAPITYNTVTPITVNP 112

Qy 49 PFT-----YKTFYVPKX-LSQKHIRLYFAAVNTDCEVFLNGEKVGENHIEYLPFEV 99

Db 113 FVPVTENPTGCGSLTENVDSEWLGCGQTRIIFDGVNSAFHLMCNGRWGYGQDSRLISEF 172

Qy 100 DVTGKVKSGENELRVVENRLKVGSPPSKVPDSCGTHTVGFGSFPRANFDFEPYGGIIRP 159

Db 173 DLSAFLRAGENRLAVMV-LRWSDGSX---LEDQ-----DMMRMSGIFRD 212

Qy 160 VLIETDহারিদিমবদতসেসেপেক্ল--GKVYKIEVSEAVGQEMTIKLGEEEKKIR 216

Db 213 VSLHKPTTQISDFHVATLFNDFSRVLAFAVQWAGELRDEL--RVTVSLWQGETVYA 269

Qy 217 TSNRFVEGEFI-----LENARFWSLEDPYLYPLKVELEKDEYTL----- 255

Db 270 SGTAPEGGEIIDERGGYADRVTLGLNVENPKLWSAETPNIRAVVELHTADGTLIEAEAC 329

Qy 256 DIGRTISWDEKRLYLNGKPVFLKFGKHEEFVLGCGTFPYPLMIKPNLLKWINANSFR 315

Db 330 DVGFREVRIENTGLLLNGKPLLRGVNRHEHHPHLCQVMDQTMQDILLMKONNFNAV 389

Qy 316 TSHYPYSEEWLIDLADRLGILVIDEA--PH-VGITRYHYNPETQKIAEDNIRRMIDRHK 370

Db 390 CSHYPNHPLMYTLCDRYGLYVDEANIETHGMVPMNRLLTDDPRWLPMASERVTRMVQDR 449

Qy 371 NHPSVIMWSVANEP--ESNHPDABGFFKALYETANEMDRTRPV-----VMVS 415

Db 450 NHPSVIIMSLNESGHGANH-----DALYRWIKSVDPSPRPVQEGGADTSATDILCP 502

Qy 416 MMDAPDERTROVALKYFDIVCVNRYGYWYIYQGRIEGLQALEKDIIEELYARHKPIFVT 475

Db 503 MYARVDEDDQFPRAVPKMSI-----KKWLSLPGEM-----RPLILC 537

Qy 476 EFG---ADAIAGIHYDPQMFSESEYA-----ELVEKTRILLKKDYIIGTH 519

Db 538 EYAHAMGNSLGG-----FAKYMQAFRQYPRLOGGFVMDVLVDQS---LIKIDE--NGN 584

Qy 520 WAAF--ADFKTPONVVRPILNHHKGVFTRDRQP 549

Db 585 PWSAYGGDFGDTPNDRQFCMN--GLVFADRTF 614

Search completed: January 24, 2005, 07:53:30  
Job time : 29 secs

Job time : 29 secs

Page Blank (uspto)